Model-based shape matching of orthopaedic implants in RSA and fluoroscopy

A.H. Prins

PhD Thesis, Leiden University Medical Center, Leiden, The Netherlands

Cover design and Layout: A.H.Prins

Printed by: CPI Koninklijke Wöhrmann

Model-based shape matching of orthopaedic implants in RSA and fluoroscopy

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof. mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op
donderdag 4 juni 2015
klokke 16:15 uur

door

Anne Hendrik Prins

geboren te Assen
in 1981

Samenstelling promotiecommissie:

Promotor:	Prof. dr. ir. E.R. Valstar
Co-promotores:	Dr. ir. B.L. Kaptein
	Dr. B.C. Stoel
Overige leden:	Prof. dr. R.G.H.H. Nelissen
	Prof. dr. ir. J.H.C. Reiber
	Prof. dr. J. Harlaar (VUMC, Amsterdam)
	Dr. S.A. Banks (University of Florida, Gainesville, US)

for Andrea

CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Radiological assessment of implant position and kinematics

In patients with severe arthritis, e.g. osteoarthritis or rheumatoid arthritis, the replacement of the degenerated joint by a prosthesis is a common and successful surgical procedure. Approximately ninety percent of the implants will function well for up to 15 years, greatly alleviating pain and improving the function of the joint [Gill et al., 1999, Kim et al., 2001, Costigan et al., 2002, Malchau et al., 2002, Banks et al., 2003, Catani et al., 2006, Havelin et al., 2009]. However, some implants fail much earlier, e.g. due to an infection or dislocation, but most commonly due to aseptic loosening. As a consequence another major surgery is eventually needed to replace the implant.

Radiolucency measured on standard clinical radiographs has been shown to be an important indicator of implant loosening. However, radiolucency is an indirect measure of possible prosthesis loosening and it can be underestimated on a radiograph or can be difficult to measure due to overprojection [Nelissen, 1995, Reading et al., 1999]. The migration of the implant over time with respect to the bone is an alternative measurement of implant loosening.

Since assessing migration from standard clinical X-rays is insufficiently accurate [Malchau et al., 1995], Roentgen stereophotogrammetric analysis (RSA) has been developed as a highly accurate alternative technique to measure implant migration [Selvik, 1989]. Small radiopaque markers are attached to the implant and the surrounding bone (Figure 1.1). The marker positions are measured with sub-millimeter accuracy by a synchronized and carefully calibrated dual X-ray setup [Mjöberg et al., 1986, Kärrholm, 1989, Kärrholm et al., 1994, Nelissen et al., 1998, Valstar, 2001, Börlin et al., 2002]. The migration is measured as the relative change in position between the implant markers and the bone markers. It has been found that this measurement is a good predictor of aseptic loosening [Kärrholm et al., 1994, Ryd et al., 1995, Pijls et al., 2012, Nieuwenhuijse et al., 2012].

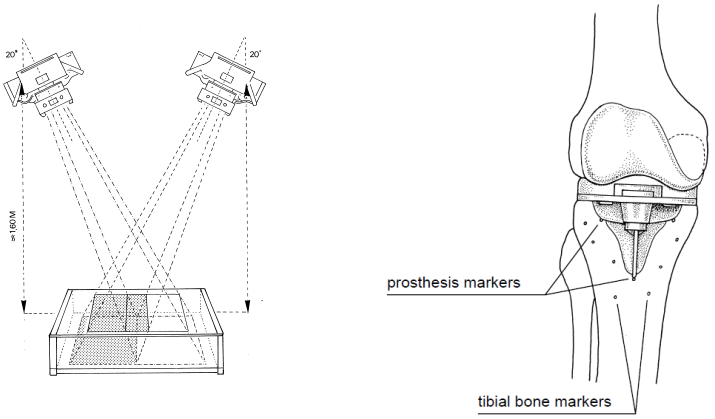


Figure 1.1: Typical RSA setup with two Roentgen foci (left) and a schematic of a knee implant with markers in the surrounding bone (right)

Where implant position is measured to assess migration, the measurement of *in vivo* implant motion is an important tool for validating the implant design by comparing actual *in vivo* implant motion with its designed movements. The external limb motion is often measured with a motion capture system which measures the positions of small markers attached to the skin with multiple camera's. However, the measurement of implant motion is inaccurate using external motion capture, because the skin and attached markers move relative to the implant [Sati et al., 1996, Leardini et al., 2005, Garling et al., 2007, Barre et al., 2013]. In addition, such external measurements cannot measure the internal kinematics of for example the mobile bearing in a knee implant [Garling et al., 2005, Wolterbeek et al., 2009, 2012a].

Fluoroscopic analysis is a technique well suited for measuring the internal joint kinematics [Banks and Hodge, 1996, Hoff et al., 1998, Zuffi et al., 1999, You et al., 2001, Komistek et al., 2003, Kanisawa et al., 2003, Tashman and Anderst, 2003, Li et al., 2004, Garling et al., 2005, Mahfouz et al., 2003, Hanson et al., 2006, Li et al., 2008]. With fluoroscopic analysis, X-ray video is used to record bone and implant motion at frame rates up to 30 Hz in clinical research. The patient can perform a (dynamic) task, while fluoroscopy captures the dynamic behavior of the implant and bone during that task [Banks and Hodge, 1996, Hoff et al., 1998, Zuffi et al., 1999, You et al., 2001, Komistek et al., 2003, Kanisawa et al., 2003, Tashman and Anderst, 2003, Li et al., 2004, Garling et al., 2005, Mahfouz et al., 2005, Hanson et al., 2006, Li et al., 2008].

1.1.1 Model-based RSA

At the Leiden University Medical Center, RSA software has been developed with a model-based approach to determine the implant position and orientation from stereo radiographs (Model-based RSA) [Valstar, 2001, Kaptein et al., 2003]. Without the need for attaching markers to the implant, model-based RSA determines the position and orientation of the implant from the shape of its silhouette in the radiographs. Markers are still necessary in the bone as model-based matching of the bone is not yet accurate enough.

The outer contour of the implant silhouette is extracted from the X-ray image and shape matching determines the position and orientation of an accurate 3D model of the implant, such that a virtual projection of a 3D model of the implant matches the implant silhouette. Computer-aided design (CAD) implant models were initially used for Model-based RSA, but the accuracy of a reverse-engineered (RE) model was shown to increase the accuracy of Model-based RSA [Kaptein et al., 2003]. In clinical practice, the relative position and orientation of the implant can be measured with errors smaller than 0.5 mm and 0.5° [Kaptein et al., 2006]. The method has been applied successfully in clinical studies to measure implant position and migration [Nelissen et al., 2005, 2002, Nieuwenhuijse et al., 2012, Pijls et al., 2012].

A possible limitation of model-based RSA is introduced, where the specific shape of the implant model makes model-based shape matching in some cases infeasible or inaccurate. Shape matching relies on a sufficiently unique silhouette such that the implant position and orientation can be determined. For example, rotating a hip stem about its longitudinal axis results in no changes or minor changes to its silhouette due to its cylindrical shape. This makes the measurement of this longitudinal rotation inaccurate.

1.1.2 Model-based fluoroscopic analysis

Model-based measurements, as used in RSA, have also been extended to fluoroscopic analysis and are applied for measuring knee implant kinemat-

ics [Garling et al., 2005, Wolterbeek et al., 2009, 2012a,b]. A few milestone papers were published halfway the nineties on measuring knee implant kinematics utilizing single-plane fluoroscopy [Stiehl et al., 1995, Banks and Hodge, 1996]. Stiehl et al. [1995] measured the knee kinematics of 47 patients by estimating the position and orientation of the implant in static radiographs every 5° of knee flexion. Banks and Hodge [1996] demonstrated the feasibility of fluoroscopy, or X-ray video, to record knee kinematics. Similar methods have been developed for measuring the position and orientation of the implant for model-based fluoroscopic analysis [Zuffi et al., 1999, Mahfouz et al., 2003, Komistek et al., 2003, Kanisawa et al., 2003, Li et al., 2004, Hermans et al., 2007, 2008].

The estimation of the position and orientation of the implant has been done using features, intensities or gradients [Markelj et al., 2012]: feature-based methods perform pose estimation on features extracted from the image such as the outer contour of the implant's silhouette. Intensity-based or gradient-based methods perform the estimation directly on the image data or after edge detection in the image.

The feature-based approach for fluoroscopic analysis has been further developed in this thesis. The features of the implant silhouette are detected in the X-ray video frame. A model-based shape matching approach, similar to the one used in model-based RSA, estimates the position and orientation of a 3D implant model. An optimization method minimizes an error measure, such that a virtual projection of a 3D model of the implant matches the implant silhouette.

The reported accuracy for fluoroscopic analysis ranges from 0.09 mm to 0.40 mm for the in-plane positions and from 0.35° to 1.3° for the orientations [Banks and Hodge, 1996, Hoff et al., 1998, Zuffi et al., 1999, Komistek et al., 2003, Kanisawa et al., 2003, Tashman and Anderst, 2003, Li et al., 2004, Garling et al., 2005, Mahfouz et al., 2005, Hanson et al., 2006, Li et al., 2008]. For most clinical research questions, an accuracy of 1 mm for the in-plane positions and 1° for the orientations would be sufficient.

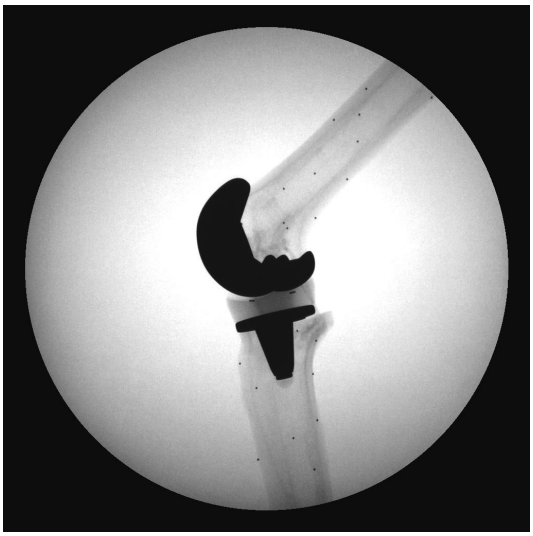


Figure 1.2: Typical high quality fluoroscopic frame of a phantom experiment

However, poor accuracy with errors of several millimeters has been reported for the out-of-plane position due to the single-plane nature of fluoroscopic analysis [Banks and Hodge, 1996, Hoff et al., 1998, Mahfouz et al., 2003, Komistek et al., 2003, Kanisawa et al., 2003]. With fluoroscopic analysis of a knee implant, this can result in the femoral component seemingly intersecting with the polyethylene insert, which is physically impossible.

As another limitation, the resolution and contrast of fluoroscopic frames are much lower compared to clinical radiographs and large image deformation is present on analogue systems with image intensifiers. When capturing a dynamic task with a high frame rate, a compromise needs to be found between exposure time, X-ray intensity, pulse width and radiation exposure for the subject. Such a compromise may result in poor image quality, which makes it difficult to distinguish the implant silhouette from the surrounding bony structures and tissues in the frame.

The analysis of fluoroscopic data is still mostly a frame-by-frame analysis with several manual or semi-automatic tasks. There is however no information available on the overall performance of these methods in terms of accuracy, robustness, computational cost, etc. In addition, the inaccuracies or the specific effects of the shape matching method on the accuracy and robustness of single-plane fluoroscopic analysis are unknown.

The labour-intensive nature makes model-based shape matching operator-dependent and possibly less robust. The number of frames to be analyzed after a single session can run into the hundreds or thousands when a captured dataset can easily encompass a few seconds of video with frame rates up to 30 Hz. This makes the analysis time-consuming and limits the method to small scale study groups.

1.2 Aim

Model-based shape matching methods for RSA and fluoroscopy are valuable for measuring implant migration and implant kinematics. However, several limitations have been described in the reliability and usability of such methods. Therefore, the aim of this research is to improve the reliability and usability of model-based shape matching for RSA and fluoroscopic analysis. Therefore, current limitations have been investigated in this thesis and new approaches have been developed:

Improvements to the interactivity of the shape matching method:

The labour-intensive nature of model-based shape matching makes the method operator-dependent and possibly less robust. The selection of relevant contour parts needs to be done manually and the researcher needs to review the results of pose estimation each frame and restart the process in case of suboptimal solutions.

A new model-based shape matching method will be presented with integrated contour detection, which improves the interactivity and ease-of-use of the algorithm, thereby making the pose measurements more robust and less operator dependent when dealing with poor image quality (Chapter 2).

Until now, there was no information available on the overall performance of these methods in terms of accuracy, robustness, computational cost, etc. Therefore, Chapter 5 will assess the performance of several optimization methods for model-based shape matching.

A combined model approach to improve the accuracy: The specific shape of the implant model could cause a failure in model-based shape matching or limit the accuracy. For example, rotating a hip stem about its longitudinal axis results in no changes or minor changes to its silhouette due to its cylindrical shape. This makes the measurement of this longitudinal rotation inaccurate.

A solution will be presented in Chapter 3 for increasing the accuracy of pose estimation for hip stems by adding the spherical head to the model with an additional degree of freedom.

In Chapter 4 the out-of-plane accuracy in single-plane fluoroscopy is improved by combining the femoral and tibial component into a single model. By adding a collision constraint, physically impossible intersections between the femoral component and the polyethylene insert are prevented.

Practical consequences of single-plane inaccuracy: The loss of contact *vivo* between the femoral and tibial components, also known as condylar lift-off, is often reported based on detailed measurements of the component positions with single-plane fluoroscopy. In Chapter 6, the practical consequences of low accuracy of single-plane fluoroscopy is assessed by comparing fluoroscopic lift-off measurements to actual *in vivo* contact force, as measured with a knee implant instrumented with force sensors.

CHAPTER 2

INTEGRATED CONTOUR DETECTION AND POSE ESTIMATION FOR FLUOROSCOPIC ANALYSIS OF KNEE IMPLANTS

A.H. Prins[1], B.L. Kaptein[1], B.C. Stoel[2], R.G.H.H. Nelissen[1], J.H.C. Reiber[2], E.R. Valstar[1,3]

1. *Biomechanics and Imaging Group, Department of Orthopaedics, Leiden University Medical Center, The Netherlands*

2. *Division of Image Processing, Department of Radiology, Leiden University Medical Center, The Netherlands*

3. *Department of Biomechanical Engineering, Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, The Netherlands*

Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine 2011 225 (8), 753–761

Abstract

With fluoroscopic analysis of knee implant kinematics the implant contour must be detected in each image frame, followed by estimation of the implant pose. With a large number of, possibly low quality, images, the contour detection is a time-consuming bottleneck. In this article an *Automated* contour detection method is proposed, which is integrated in the pose estimation.

In a phantom experiment the *Automated* method was compared to a *Standard* method, which uses manual selection of correct contour parts. Both methods demonstrated comparable precision, with a minor difference in the Y-position (0.08 mm vs. 0.06 mm). The precision of each method was so small (below 0.2 mm and 0.3°) that both are sufficiently accurate for clinical research purposes.

The efficiency of both methods was assessed on six clinical datasets. With the *Automated* method the observer spent 1.5 minutes per image, significantly less than 3.9 minutes with the *Standard* method. A Bland-Altman analysis between the methods demonstrated no discernable trends in the relative femoral poses.

The threefold increase in efficiency demonstrates that a pose estimation approach with integrated contour detection is more intuitive than a *Standard* method. It eliminates most of the manual work in fluoroscopic analysis, with sufficient precision for clinical research purposes.

2.1 Introduction

Single-plane fluoroscopic analysis is an important tool for the evaluation of knee implant kinematics. Many methods have been described for the estimation of the three-dimensional (3D) position and orientation (pose) of the implant in each fluoroscopic image. Template-matching [Banks and Hodge, 1996, Hoff et al., 1998] or model-based 3D-to-2D registration [Zuffi et al., 1999, Kaptein et al., 2003] are common approaches. These methods have accuracies ranging from 0.09 mm to 0.40 mm for the in-plane positions and from 0.35° to 1.3° for the orientations [Banks and Hodge, 1996, Hoff et al., 1998, Mahfouz et al., 2005, Komistek et al., 2003, Kanisawa et al., 2003, Li et al., 2004, Garling et al., 2005, Hanson et al., 2006]. These in-plane accuracies are sufficient for clinical uses, whereas the out-of-plane position is considered not accurate enough for usage in many clinical applications.

Many of these methods require that the contour of the implant is detected in each image frame. The implant pose is then estimated by minimizing the difference between the contour and a virtual projection of the model. Contour detection is often a manual or semi-automatic task, which requires a significant amount of user interaction: selecting the relevant contour parts or discarding the erroneous parts. Since fluoroscopic images have lower image contrast and resolution than standard X-rays and the contour detection must be performed for each single image in a dataset, this makes the analysis cumbersome and time-consuming. In addition, the accuracy of the detected contour is an important factor in the final accuracy of the estimated pose [Fregly et al., 2005, Mahfouz et al., 2005]. Because Mahfouz et al. [2005] considered contour detection too prone to errors, they suggested to avoid an a priori contour detection step and instead use a direct model-to-image pose estimation method.

The goal of this study was to validate a new and automated model-based contour detection method, which is integrated into a model-based pose estimation method. If the automatic contour detection turns out to be of sufficient accuracy and reproducibility, the method will be much more intuitive and efficient to use for the researcher. The analysis of a complete fluoroscopic dataset can easily be automated by propagating the pose from

one image to the next [Zuffi et al., 1999].

Both phantom and clinical data were used to validate the accuracy and precision on the clinically relevant in-plane positions and orientations. It is compared to a conventional model-based pose estimation method [Kaptein et al., 2003] with semi-automatic contour detection (the Canny edge detector [Canny, 1986]), and the effects of image quality, the agreement in pose and the analysis time of the methods were investigated as well.

2.2 Method

The main input of the automated model-based contour detection method consists of a fluoroscopic image, the relative X-ray focus position and a 3D surface model of the implant. Furthermore, an initial candidate pose is required.

The image is preprocessed by applying noise reduction with a Gaussian filter. The integrated pose estimation and contour detection consist of the following loop (Algorithm 1), updating the contour and pose in each iteration:

Algorithm 1 Pose estimation with integrated contour detection

1: **for each** iteration **do**
2: **Detection:**
Model-based contour detection, based on the current candidate pose.
3: **Selection:**
Contour-point selection, automatically selecting 20% of the (good-quality) contour parts.
4: **Pose estimation:**
Robust pose estimation with the selected contour parts, giving a new candidate pose.

Each of these three steps is described in more detail below. A small number of iterations (typically five) with the above steps is often sufficient for the method to converge to the desired pose and contour. As a final post-processing step, the same robust pose estimation as in iteration-step 3 is applied to the final contour, but using all the good-quality contour parts. This makes pose estimation slower, but also more accurate.

The initial pose for the method can be provided by the user. For example, in our application the user can easily and intuitively manipulate the 3D surface model and can get direct feedback of the implant pose with respect to the image. A second possibility is the propagation of the implant pose from the analysis of an earlier image frame [Zuffi et al., 1999], and this allows for the automatic analysis over all image frames.

2.2.1 Model-based contour detection

First, a virtual projection of the model onto the image plane is calculated (Figure 2.1a) [Kaptein et al., 2003]. This projection is a closed curve and represents the outer boundary of the implant silhouette.

A region of the image around the virtual projection is then resampled along scan lines perpendicular to the virtual projection (counterclockwise and from outside to inside). The resulting scan-matrix represents a straightened version of the image region in a band around the virtual projection. Each line in the matrix corresponds to a scan line starting at the outside of the virtual projection and ending on the inside of the virtual projection (Figure 2.1b). The width of this region can be adjusted by setting the length of the scan lines; the pixel spacing is the same as in the original image.

A derivative matrix is calculated with a convolution operation ($[1\ 0\ -1]$ kernel) along each line in the intensity matrix. This represents the derivative in the image perpendicular to the virtual projection. With a dark implant silhouette, the positive edges (from black to white) will be assigned a high value, while negative edges (from white to black) a low value (Figure 2.1c).

A dynamic programming approach extracts an optimal path, passing each scan line once, while maximizing the sum of edge values in the edge matrix [Bellman and Dreyfus, 1966]. The resulting path follows decreasing edges (from white to black) as close as possible (Figure 2.1c). This path is transformed back into the image domain (Figure 2.1d) and for each point the edge strength (derivative edge value) is stored.

In the original image domain the result is a new contour within the band

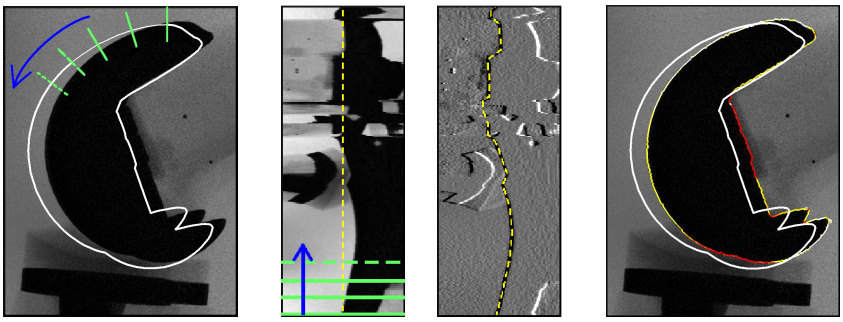


Figure 2.1: ***a)*** *Contour detection starts with a virtual projection.* ***b)*** *The image is resampled (counter-clockwise) along this virtual projection with scan-lines perpendicular to the projection, resulting in a scan-image with the virtual projection as a straight line in the center (dashed line) and the corresponding scan lines.* ***c)*** *A convolution with a positive difference filter ([-1 0 1]) results in the cost-image (third image). Minimal path extraction using dynamic programming then extracts a path (dashed curve).* ***d)*** *After back-transforming this path, the result is a contour (right image) with associated edge strengths.*

around the virtual projection, with the edge strengths available for later use by the pose estimation.

2.2.2 Contour-point selection

The edge strengths are normalized over the entire contour to a range of $[0, 1]$. Using a threshold in the range of $[0, 1]$, the user can specify how much of the contour should be discarded. Contour points with an edge strength below the threshold are then discarded (a threshold of 0.25 was typically used). From the remaining contour-points (those with an edge strength above the threshold), 20% was selected, uniformly sampled along the contour, in order to reduce the computational cost.

2.2.3 Robust pose estimation

A global optimization method (Down-Hill-Simplex method [Nelder and Mead, 1965] combined with Simulated Annealing [Kirkpatrick, 1984]) finds the pose, for which the virtual projection fits the selected contour points best.

To this end, a weighted distance measure is calculated between the selected contour points and the virtual projection of the implant model. The distance measure is the same as described by Kaptein et al. [Kaptein et al., 2003], with one addition: The edge strengths from the selected contour points are used as weights.

2.3 Validation: experimental

Data was collected from a phantom experiment using a bi-plane flat panel fluoroscopic setup (Super Digital Fluoroscopy (SDF) system, Toshiba Infinix: Toshiba Medical Systems Europe, Zoetermeer, The Netherlands). The phantom study was performed with a size 3 cruciate-substituting PFC-Sigma prosthesis fixed in sawbones with a 5 mm thick insert (DePuy Orthopedics, Warsaw, IN). A 3D implant model of the femoral component was reverse engineered with an accuracy of 0.05 mm (TNO Industry, Eindhoven, The Netherlands) for use by the pose estimation methods.

The image intensifiers were positioned perpendicular to each other and the sawbones were placed such that one image intensifier had a medial-lateral view, while the other had an anterior-posterior view. The X-ray focus positions were calculated using a calibration box [Koning et al., 2007].

Two motions of the femur were captured (15 fps). In the first motion, the femur moved from full extension to 90° of flexion, followed by an abduction of approximately 20°, back to 20° adduction and finally back to full extension. In the second motion, the femur started at 30° of flexion, moved to full extension after which some internal/external rotation (roughly 20°) was performed.

A subset of the images ($N = 58$) was used for the experiments capturing a broad range of poses. A standard model-based pose estimation method (Model-Based RSA 3.21, Medis Specials, Leiden, the Netherlands [Kaptein et al., 2003]) was applied to the images from both image intensifiers and this was used as the *Reference* measurement.

2.3.1 Image quality

The new method was validated on single-plane fluoroscopic image data using the data from the phantom experiment, but only from the image intensifier with a medial-lateral view. This data was of excellent quality; high resolution and high image contrast. In routine clinical practice, the image quality is worse and can have an influence on the accuracy of contour detection and pose estimation. In this validation experiment the effects of lower image quality on the new method was investigated.

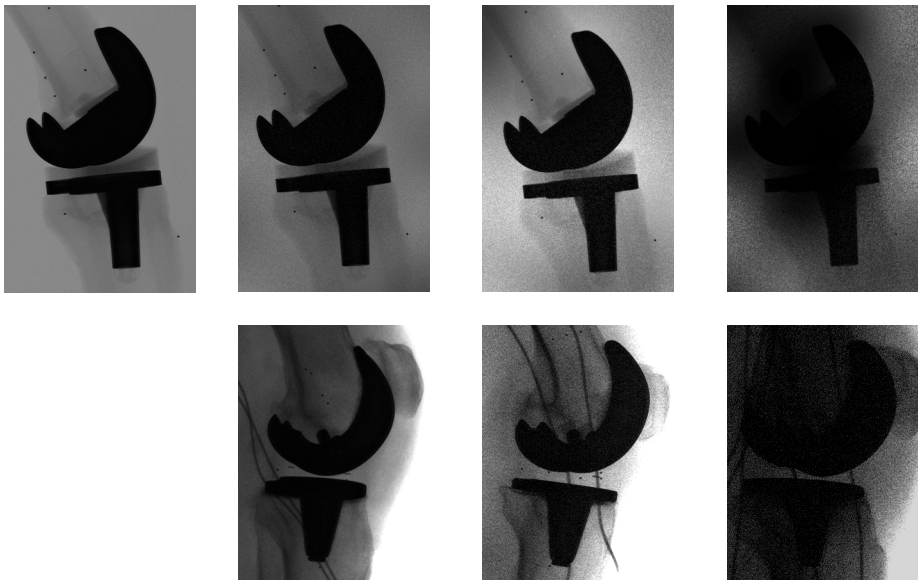
The image qualities were assessed as encountered in ongoing clinical studies. Four quality levels were defined: *L0*, *L1*, *L2* and *L3*. Level *L0* has no quality reduction and serves as a baseline measurement. Levels *L1* - *L3* represent good, moderate and poor qualities, respectively. Each image was degraded according to these levels with the following method:

1. A template image was created with Simplex noise [Perlin, 2001] which was used as a template for local intensity-reduction. This resulted in different parts of the image having different intensity levels, which was used to crudely simulate different contrasts, which can arise in clinical practice due to soft tissue.

2. Locally varying Gaussian noise was introduced and the overall sharpness of the image was reduced with a Gaussian blur.

The parameters used, the example images and the corresponding clinical images are presented in Figure 2.2

The new *Automated* method was applied to all the images with an experienced user providing the initial pose of the model. For comparison, a standard model-based pose estimation method (Model-Based RSA 3.21, Medis Specials [Kaptein et al., 2003]) was also applied to the images. In this *Standard* method an experienced user was responsible for supervising the Canny edge detection by selecting and cutting out contour parts and for providing an initial pose. The pose was then estimated with iterative-inverse perspective matching [Wunsch and Hirzinger, 1996] followed by a global optimiza-



Original image No modifications	No blur Noise variance: 0.05 Grey values: $[0.2, 0.6]$	Blur radius 1px Noise variance: 0.1 Grey values: $[0.1, 0.4]$	Blur radius 1px Noise variance: 0.2 Grey values: $[0.0, 0.3]$

Figure 2.2: Examples of reduced quality images of the phantom experiment in the first row and their clinical counterparts on the second row. The intensity-range on a scale of $[0, 1]$ *is presented for each image in the third row, together with the variance for the noise.*

tion of the distance measure between the contour points and the implant model [Kaptein et al., 2003].

For both the *Automated* and the *Standard* method, the error in pose was calculated as the difference for each pose-parameter with respect to the pose obtained by the bi-plane *Reference* measurement. Students T-test was then used to compare the mean errors between the *Standard* and the *Automated* and Levene's test to compare the variance of the differences.

2.4 Validation: clinical

Six fluoroscopic datasets were collected from two patients with a ROCC prosthesis (Biomet, Warsaw, IN) with a total of 266 images. These datasets were acquired with single-plane flat-panel fluoroscopy (15 fps) while these patients were performing a step-up task. The three datasets of one patient

were considered of moderate quality, while those of the other patient were considered of good quality, comparable to levels *L1* and *L2* in Figure 2.2.

An experienced user applied both the *Automated* and *Standard* methods to each dataset and the total analysis time was recorded for each method. The tibial and femoral components were analyzed and the relative poses between the two components were calculated for each method. To determine the agreement between the two methods, a Bland-Altman analysis was performed for each pose parameter.

2.5 Results: experimental

All the errors in pose for both methods are presented in Table 2.1. The mean errors and the standard deviations for the highest and lowest image qualities (*L0* and *L3*) are presented in Figure 2.3. The most prominent differences are in the systematic errors between the two methods: The *Automated* method demonstrates small, but statistically significant worse systematic errors in the in-plane positions (X: 0.07 mm, Y: 0.08 mm), which is consistent over all quality levels. The mean error in the less accurate out-of-plane position increases up to 0.77 mm for the *Automated* method and is significantly larger ($p < 0.001$) than the *Standard* method for all quality levels.

The standard deviations are very well comparable between the two methods and with values below 0.1 mm and 0.1° sufficiently precise for most clinical uses. From quality levels *L1–L3*, the standard deviation in the Y-position (0.06 mm) of the *Standard* method is slightly better ($p < 0.02$) than that of the *Automated* method (0.07 mm on *L0* up to 0.08 mm on *L3*). In turn the *Automated* method demonstrates a smaller ($p = 0.03$) standard deviation on the X-orientation at *L0* (0.07° vs. 0.09°) and a higher ($p < 0.02$) standard deviation in the Y-orientation (0.13° vs. 0.09°) and Z-orientation (0.3° vs. 0.21°). When considering the magnitude of these standard deviations, the two methods perform very similar.

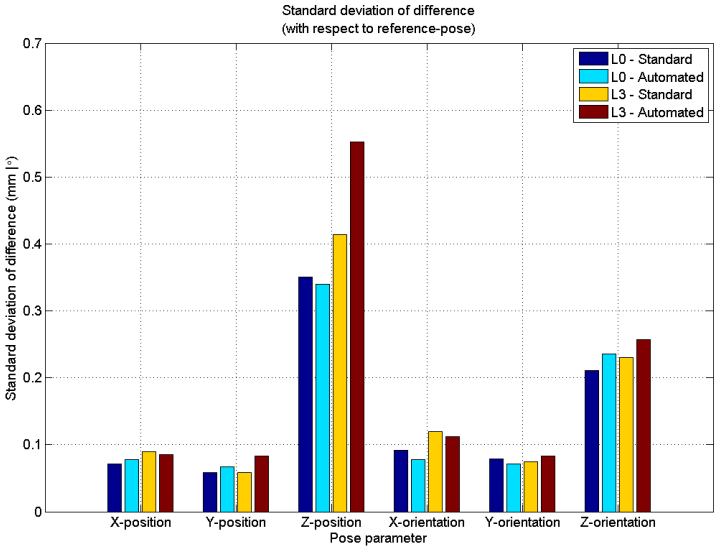


Figure 2.3: Standard deviations of the error in pose with respect to the reference pose measurement for the Automated and Standard methods and the best and the worst level of image quality.

2.6 Results: clinical

On average an experienced user spent 3.9 ± 0.36 minutes per image with the *Standard* method, whereas the *Automated* required only 1.5 ± 0.36 minutes per image. This difference was highly significant ($p < 0.001$).

The Bland-Altman plots of the relative pose with respect to the tibia pose are presented in Figure 2.4. There were no strong trends discernable in the Bland-Altman plots. The largest differences presented themselves in the out-of-plane position with a mean difference of 0.77 mm and large standard deviation of 3.46 mm. A relatively large discrepancy in the Y- and Z-orientation was found with standard deviations of 1.11° and 0.98°.

2.7 Discussion

In this study it was demonstrated that the contour detection can be completely integrated within the pose estimation. The integration of automatic

	Position						Rotation					
Method	**X**: mean	std	**Y**: mean	std	**Z**: mean	std	**X**: mean	stdev	**Y**: mean	stdev	**Z**: mean	stdev
L0: Standard	-0.03[1]	0.07	-0.02[1]	0.06	0.29[1]	0.35	0.02	0.09	-0.02[1]	0.08	-0.06	0.21
Automated	0.07	0.08	0.08	0.07	0.54	0.34	0.03	0.08	-0.08	0.07	-0.06	0.24
L1: Standard	-0.03[1]	0.08	-0.02[1]	0.05	0.11[1]	0.44	0.03	0.09[1]	-0.02[1]	0.07	-0.05	0.22
Automated	0.06	0.08	0.09	0.07	0.52	0.34	0.03	0.07	-0.08	0.09	-0.07	0.24
L2: Standard	-0.03[1]	0.08	-0.01[1]	0.05	0.15[1]	0.47	0.02	0.11	-0.04[1]	0.07	-0.06	0.23
Automated	0.06	0.08	0.08	0.08	0.58	0.4	0.02	0.08	-0.1	0.08	-0.06	0.26
L3: Standard	-0.03[1]	0.09	-0.01[1]	0.06	0.20[1]	0.41	0.03	0.12	-0.04	0.07	-0.07	0.23
Automated	0.06	0.09	0.08	0.08	0.77	0.55	0.03	0.11	-0.11	0.08	-0.06	0.26

[1]Significant difference ($p < 0.05$) between the Automated and the Standard methods on the same level of image quality.

Table 2.1: Means and standard deviations of the errors in position and orientation for both the Automated and Standard methods on single-plane data, with respect to the Reference method (from bi-plane data). The single-plane data was reduced according to four quality levels.

contour detection enables a more intuitive and thus faster analysis procedure for the researcher, with only minor consequences for the accuracy of the system. Although the systematic errors of the *Automated* method are consistently higher than the errors of the *Standard* method, their values in the clinically relevant parameters remain below 0.1 mm and 0.1°. In addition, the systematic errors are less important than the standard deviations when investigating implant kinematics, where the relative motions of the components are considered. With respect to the standard deviations, the two methods performed virtually identical. The only significant difference (0.08 mm vs. 0.06 mm) was found in the Y-position. These differences are, however, clinically irrelevant.

The systematic errors of both methods (up to 0.08 mm) in the in-plane positions are hard to explain. On quality level *L0*, the images to which the *Standard* and *Automated* methods were applied are identical to the images used by the *Reference* method. The *Standard* and *Reference* method employ the same technique, except for the *Reference* method which has a second image available from the bi-plane setup. It may be that the contour detected in a single image is slightly off with respect to the actual silhouette of the implant. With a pixel size of 0.39 mm, a difference of e.g. half a pixel could already introduce a systematic shift of 0.1 mm in the pose. Whereas the bi-plane *Reference* method can correct for such discrepancies by using the data in the second image, this is not possible with the single-plane *Standard* and

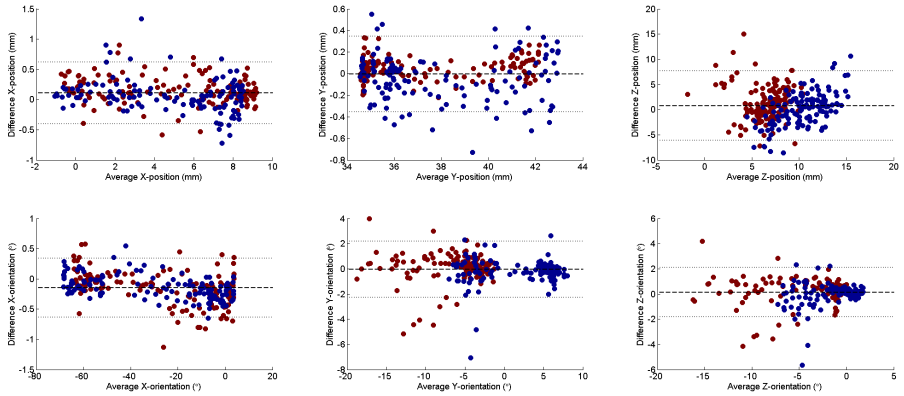


Figure 2.4: Bland Altmann plots for the agreement in the pose measured by the Standard method and the Automated method on the clinical data. The x-axes represent the agreement in the pose parameters, calculated as the mean of the two methods. The y-axes represent the difference in pose-parameter between the two methods. The colors represent the two patients from which the data was obtained. The central dashed line represents the mean difference over all measurements, while the two outer dashed lines indicate limits of agreement for the differences (two standard deviations around the mean).

Automated methods. The difference in contour detection method between the *Standard* and *Automated* methods may then cause the discrepancy in the systematic errors between the *Standard* and *Automated* methods.

The results from the phantom experiment compare well with the results published in the literature and are more accurate than the literature. Earlier studies on single-plane fluoroscopy have reported precisions for estimating positions from 0.09 mm to 0.46 mm and precisions for estimating the orientations between 0.35° and 1.3° [Banks and Hodge, 1996, Hoff et al., 1998, Mahfouz et al., 2005, Komistek et al., 2003, Kanisawa et al., 2003]. Fregly et al. (2005) have demonstrated the effects of X-ray attenuation on the accuracy of the silhouette in the image and how this could possibly account for most of measurement bias. Mahfouz et al. (2005) demonstrated how manual segmentation can severely affect the accuracy of pose estimation and even recommended a pose estimation process without manual, a priori segmentation. The *Automated* method has no such "manual, a priori segmentation", but instead relies on an automatic model-based segmentation method.

In the clinical experiments a threefold improvement was found in analysis-

time with the *Automated* method with the added advantage that no continuous supervision by the researcher was required. With a fluoroscopic dataset of 50 images, a researcher would spend 75 minutes with the new *Automated* method, and 195 minutes with the *Standard* method.

The lack of a good reference measurement in the clinical data makes it difficult to determine the accuracy of the two methods. The two methods show differences within 1 mm and 1° for all orientations and positions, respectively, except for the out-of-plane position where the differences can be as high as 7 mm. These values are comparable with the earlier mentioned ranges from the literature.

A difficult problem arose with the tibial orientation in three of the six datasets. For several images, there was some ambiguity in the silhouette of the tibial component. Its silhouette was such that both pose estimation methods had to choose from two or sometimes three pose candidates. In those cases it could be that the *Standard* and *Automated* picked different pose candidates, resulting in differences in measured orientations of a few degrees. This is likely an issue with any pose estimation method and could have caused the relatively large differences in Y-orientation and Z-orientation.

When dealing with patient-data, the motion between two consecutive images can be large. Extrapolation of the pose from one image to the next [Zuffi et al., 1999] can then put the implant model too far away from its new silhouette. This can cause the automatic model-based contour detection to fail, when the detection region on the image around the virtual projection of the model is too small to capture enough parts of the new contour. To overcome this problem, the detection region can be enlarged. However, this adds the risk that another strong contour instead of the implant contour is found, e.g. the skin-to-air transition. In those cases it was easier for the user to update the model pose interactively and then continue the analysis.

Overall, a pose estimation approach was demonstrated with integrated contour detection, which yielded precise poses in the clinically relevant directions (below 0.2 mm and 0.3°). The easier workflow, the omission of a semi-automatic contour detection eliminates most of the manual work in fluoroscopic analysis. This makes the method more objective and results in

a threefold reduction in analysis time.

2.7.1 Acknowledgement

This project was sponsored by the European Community Project DeSSOS IST-2004-27252.

CHAPTER 3

HANDLING MODULAR HIP IMPLANTS IN MODEL-BASED RSA: COMBINED STEM-HEAD MODELS

A.H. Prins[1], B.L. Kaptein[1], B.C. Stoel[2], R.G.H.H. Nelissen[1], J.H.C. Reiber[2], E.R. Valstar[1,3]
1. *Biomechanics and Imaging Group, Department of Orthopaedics, Leiden University Medical Center, The Netherlands*
2. *Division of Image Processing, Department of Radiology, Leiden University Medical Center, The Netherlands*
3. *Department of Biomechanical Engineering, Faculty of Mechanical, Maritime, and Materials Engineering, Delft University of Technology, The Netherlands*

Journal of Biomechanics 2008 41 (14), 2912–2917

Abstract

Migration measurements of hip prostheses using marker-based Roentgen Stereophotogrammetric Analysis (RSA) require the attachment of markers to the prostheses. The model-based approach, which does not require these markers, is, however, less precise. One of the reasons may be the fact that the spherical head has not been modelled. Therefore, we added a 3D surface model of the spherical head and estimated the position and orientation of the combined stem-head model. The new method using a combined stem-head model was compared in a phantom study on five prostheses (of different types) and in a clinical study using double examinations of implanted hip prostheses, with two existing methods: a standard model-based approach and one using elementary geometrical shapes. The combined model showed the highest precision for the rotation about the longitudinal axis in the phantom experiments. With a standard deviation of 0.69° it showed a significant improvement ($p = 0.02$) over the model-based approach (0.96°) on the phantom data, but no improvement on the clinical data. Overall, the use of elementary geometrical shapes was worse with respect to the model-based approach, with a standard deviation of 1.02° on the phantom data and 0.79° on the clinical data. This decrease in precision was significant ($p < 0.01$) on the clinical data. With relatively small differences in the other migration directions, these results demonstrate that the new method with a combined stem-head model can be a useful alternative to the standard model-based approach.

3.1 Introduction

Roentgen Stereophotogrammetric Analysis (RSA) is a well-known method for measuring micromotion of joint replacement prostheses and can be used to detect prosthesis loosening [Kärrholm et al., 1994]. It is used to measure the position and orientation of attached prosthesis-markers with respect to markers inserted into the bone [Selvik, 1989]. This marker-based approach is very accurate, with standard deviations ranging from 0.03 mm to 0.35 mm for translations [Mjöberg et al., 1986, Kärrholm, 1989, Kärrholm et al., 1994] and from 0.05° to 0.58° for rotations [Kärrholm, 1989, Börlin et al., 2002]. But, the marker-based approach has the disadvantage that the implant may obscure these markers in the radiograph making pose estimation impossible. Furthermore, it is expensive to attach the markers to the implant.

To prevent the requirement of attaching markers to the prosthesis, elementary geometrical shape models (EGS) can be used to determine the position and orientation of a prosthesis. For example, the center of a sphere can be used to measure the position of the spherical head [Baldursson et al., 1979, Kärrholm, 1989, Önsten et al., 1995, Kärrholm et al., 1997], while the position and orientation of a cylinder or a cone can be used to measure the position and orientation of the stem of the implant [Valstar, 1996, Valstar et al., 2001, Kaptein et al., 2006].

The method cannot be applied, however, if these elementary geometrical shape models do not fit the implant properly. Therefore, a so-called model-based approach was developed to overcome this problem. It uses a 3D surface model of the stem of the prosthesis to measure its position and orientation with respect to the bone [Valstar, 2001, Kaptein et al., 2003].

In a comparison-study [Kaptein et al., 2006] the precision of three RSA methods (marker-based, EGS-based and model-based) has been assessed using the Mallory/Head prosthesis (Biomet, Inc, Warsaw, IN). The analysis showed that both the model-based approach and EGS-based approach were not as precise as the original marker-based approach. In general, the rotation about the longitudinal axis, also known as internal rotation, and the subsidence (translation along the longitudinal axis) are the two most impor-

tant early indicators of loosening [Kärrholm, 1989, Nistor et al., 1991, Gill et al., 2002] and precise measurements of these two migration directions are important. Unfortunately, both elementary geometrical shapes and model-based RSA show relatively large standard deviations for the rotation about the longitudinal axis.

The lower precision of the model-based approach may be explained partially by the fact that it uses a 3D surface model of the stem only, without the spherical head. This makes the 3D surface model relatively symmetric about its longitudinal axis, making estimation of rotation about the longitudinal axis difficult. The head was not included in the 3D surface model, however, because dimensional tolerances in the manufacturing process influence the exact position of the head with respect to the stem.

In this paper, a new method is proposed that models the dimensional tolerances by adding a spherical head to the 3D surface model in such a way that the relative position of the head is optimised during the estimation of position and orientation of the prosthesis. This method with a combined stem-head model (CM-RSA) was validated in a phantom study and a clinical study using double examinations of implanted hip prostheses. In this study, the new method (CM-RSA) was compared with the standard model-based approach (MB-RSA) and the method using elementary geometrical shapes (EGS-RSA).

3.2 Methods

To estimate the position and orientation of the prosthesis, the following three methods were used.

3.2.1 Elementary Geometrical Shapes

This method (EGS-RSA) uses elementary geometrical shapes, identified by the user, to estimate the position and orientation of a small number of landmarks [Kaptein et al., 2006] as illustrated in Figure 3.1a. Contours are

detected in the roentgen images using the Canny Edge detector [Canny, 1986]. Shapes corresponding to the spherical head of the prosthesis are identified by the user and the position of the spherical head is computed from these shapes and the two focus positions. The mid-point of the shortest line-segment connecting these two projection lines forms the estimation of the center of the spherical head, i.e. the first landmark.

The left side of the distal part of the stem is estimated as follows. The user identifies the corresponding lines in the two roentgen images. Two planes are formed from these images using the two focus positions. The crossing line of these two planes forms the left side (in 3D) of the stem. The right side of the distal part of the stem is estimated similarly. These two 3D lines are then used to compute a central axis through the distal part of the stem.

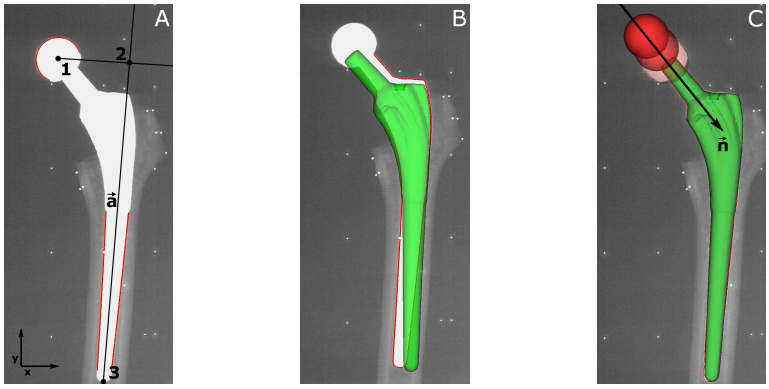


Figure 3.1: **EGS-RSA:** *The first landmark (1) is formed by the estimation of the center of the spherical head. An axis a is estimated through the most distal part of the stem. The second landmark (2) is obtained by projecting landmark 1 onto the axis a, while the third landmark (3) is formed as the projection of the tip of the stem onto axis a.* **MB-RSA:** *The 3D surface model of the stem is used to minimize the distance between a virtually projected contour and the detected contour.* **CM-RSA:** *A combined head-stem model is used to minimize the distance between a virtually projected contour and the detected contour, while allowing the position of the spherical head to vary during the minimization.*

The second landmark is obtained by projecting the first landmark, the center of the spherical head, onto this central axis.

The third landmark is defined by the most distal tip of the stem. The user identifies an initial guess of the tip in the two roentgen images. The corresponding 3D point is computed as the mid-point of the two projection lines

through the focus positions and the initial guesses of the tip. The projection of the 3D tip position onto the central axis of the stem results in the third landmark. With three landmarks, sufficient information is available to compute the position and orientation of the prosthesis in 3D.

3.2.2 Model-based RSA

This method (MB-RSA) uses a 3D surface model of the stem. This 3D surface model is used to determine the position and orientation of the prosthesis, by aligning it with the detected contours of the prosthesis in the image (Figure 3.1b): Contours are detected using the Canny Edge detector [Canny, 1986], after which the contour parts of the stem are manually selected. To reduce the computation time, 25% of the detected contour is used for the alignment of the surface model. The alignment of the surface model is performed by calculating a contour of the virtually projected 3D surface model, followed by a calculation of the distance between this virtual contour and the actual, detected, contour. The correct pose of the surface model is then determined by searching through the six-dimensional parameter (position + orientation) space for an optimal pose, which minimizes the distance between the contours.

3.2.3 Combined stem-head model

In this new method (CM-RSA), the reverse-engineered 3D surface model of the stem is combined with a 3D triangulated surface model of a sphere with a diameter of 28 mm. The vertices of the triangulated surface model of the stem corresponding to the tapered neck are manually selected and a cone is matched in a least squares manner through these vertices. The sphere model is allowed to move along the axis of this cone, resulting in a 3D surface model with seven degrees of freedom: six parameters describing the general position and orientation of the stem and an additional seventh parameter describing the position of the sphere on the estimated axis (Figure 3.1c). The default position of the sphere is specified manually by setting a suitable initial value for this seventh parameter.

Contours are detected using the Canny Edge detector [Canny, 1986], after which the correct contour parts are manually selected. Because the 3D surface model is now composed of a stem and a head, contour parts for the head are selected as well. To reduce the amount of computation time, 25% of this detected contour is used for the aligning of the surface model. Equivalent to the MB-RSA method, the alignment of the surface model is performed by calculating contours of the virtually projected combined model, followed by a calculation of the distance between this virtual contour and the actual, detected, contour. The correct pose is then determined by searching through the seven-dimensional parameter space for an optimal pose which minimizes the distance between these contours.

3.3 Experimental setup

The new method (CM-RSA) was validated in both a phantom study and on clinical data, where it is compared to MB-RSA and EGS-RSA.

3.3.1 Phantom experiment

The phantom study was performed on five prostheses (of three different hip stem designs).

- Mallory/Head size 9 and size 12 (Biomet, Inc, Warsaw, IN, USA)
- SL-Plus size 4 (Plus Orthopedics AG, Rotkreuz, Switzerland)
- Two straight Stanmore prostheses size 3 (Biomet, Inc, Warsaw, IN, USA)

A spherical head with a standard diameter of 28 mm was attached to each prosthesis and each prosthesis was rigidly fixed in a sawbone. Between 5 and 8 marker beads were inserted into each sawbone, resulting in a distribution of the markers with conditions number varying between 13 and 19 [Söderkvist

and Wedin, 1993]. For the SL-Plus prosthesis, a cup was added in which the spherical head was placed.

Eleven RSA-images were made of each prosthesis. Between exposures, the entire phantom was placed in a position and orientation mimicking the clinical situation. Furthermore, between exposures, the phantom was repositioned manually, each time in a similar clinically relevant position, but with an overall variation of roughly 15° orientation. The RSA-setup consisted of two synchronized roentgen tubes at 1.5 m above the roentgen film and were directed towards the roentgen film, each one making an angle of 20° with the vector perpendicular to the roentgen film. Migrations between pairs of consecutive scenes were computed. I.e. the migrations were computed between the following scene-pairs: (1, 2), (2, 3), (3, 4).

3.3.2 Clinical data

In addition to the phantom experiment, the CM-RSA method was applied to clinical double-examination data. 11 double stereo roentgen-images were available for the Mallory/Head prostheses (sizes $7-14$) with a metal-backed cup. For the standard Stanmore-prostheses (size 2 and size 3), eleven double images were available with a polyethylene cup. In both cases, migrations were computed between corresponding image pairs.

3.3.3 Validation

Laser-scanning (TNO Industry, Eindhoven, The Netherlands) with a spatial resolution of 0.05 mm, was used to generate reverse-engineered 3D surface models for the prostheses. The images were analyzed using Model-based RSA 3.12 (Medis specials, Leiden, The Netherlands) using standard methods for calibration and the described three methods (MB-RSA, EGS-RSA and CM-RSA) for the estimation of the position and orientation of the prosthesis.

Between exposures a prosthesis is considered to be rigidly fixed and no migration is expected of the prosthesis with respect to the tantalum markers

in the (saw)bone. Therefore, the measured migration represents the measurement error. The mean migrations give an indication of the systematic error of the measurements, while the standard deviations give an indication of the precision of the measurements. Migrations were computed using the calibration box as the global coordinate system: the reference coordinate system is illustrated in Figure 3.1a, with the x- and y-axes in the image plane and with the z-axis perpendicular to the image plane.

Levene's test for equality of variances was used to compare the standard deviations of the three methods.

3.4 Results

The time needed for the analysis of a pair of roentgen-images was comparable between the three methods and ranged from three to five minutes. For each method and each prosthesis, means and standard deviations for both the translational and rotational components of the measured migrations were computed. The results from the phantom-experiment and the clinical double-examinations are presented in Table 3.1.

Method		**Translations**			**Rotations**		
		x (mm)	y (mm)	z (mm)	x $(^\circ)$	y $(^\circ)$	z $(^\circ)$
Phantom	EGS-RSA	0.02 [0.05]	0.00 [0.10]	0.00 [0.12]	0.01 [0.15]	0.01 [0.51][1]	−0.01 [0.09]
Mallory-Head	MB-RSA	−0.01 [0.05]	0.00 [0.06]	0.00 [0.13]	0.00 [0.22]	0.00 [1.01]	0.00 [0.13]
($N = 22$)	CM-RSA	−0.01 [0.05]	0.00 [0.06]	0.00 [0.09]	0.00 [0.22]	0.00 [0.67]	0.01 [0.12]
Phantom SL	EGS-RSA	−0.02 [0.07]	0.00 [0.13]	0.01 [0.29][1]	−0.09 [0.34]	−0.01 [2.22][1]	−0.00 [0.20][1]
SL_Plus	MB-RSA	−0.01 [0.04]	0.01 [0.07]	0.00 [0.10]	−0.04 [0.27]	0.00 [1.00]	−0.00 [0.09]
($N = 9$)	CM-RSA	0.00 [0.07]	0.01 [0.11]	−0.01 [0.49][1]	−0.04 [0.22]	0.00 [1.23]	−0.00 [0.09]
Phantom	EGS-RSA	0.00 [0.04][1]	0.00 [0.09][1]	0.01 [0.23][1]	0.01 [0.15]	0.00 [0.66]	0.02 [0.08]
Stanmore	MB-RSA	0.00 [0.02]	0.00 [0.03]	0.01 [0.09]	−0.02 [0.15]	0.01 [0.94]	0.04 [0.11]
($N = 22$)	CM-RSA	0.00 [0.03][1]	0.00 [0.02]	0.00 [0.06]	0.00 [0.13]	0.00 [0.37][1]	0.01 [0.04][1]
Clinical	EGS-RSA	−0.02 [0.12]	0.11 [0.21]	−0.12 [0.30]	0.03 [0.27]	−0.41 [1.02]	0.06 [0.08]
Mallory-Head	MB-RSA	0.00 [0.08]	−0.02 [0.11]	0.02 [0.17]	0.02 [0.22]	−0.02 [0.52]	0.03 [0.10]
($N = 11$)	CM-RSA	−0.05 [0.14]	−0.03 [0.19]	0.10 [0.20]	0.02 [0.20]	−0.07 [0.43]	0.04 [0.09]
Clinical	EGS-RSA	0.01 [0.03]	0.04 [0.06]	0.01 [0.08]	0.01 [0.15]	0.12 [0.32]	−0.02 [0.06]
Stanmore	MB-RSA	0.02 [0.05]	0.01 [0.07]	0.00 [0.08]	0.00 [0.18]	0.20 [0.24]	−0.03 [0.05]
($N = 11$)	CM-RSA	0.03 [0.06]	0.02 [0.07]	−0.04 [0.11]	−0.02 [0.16]	0.17 [0.20]	−0.02 [0.06]

[1]Significantly different from MB-RSA

Table 3.1: Migration results (mean [standard deviation]) for each method applied to each prosthesis.

For the two most important indicators of prosthesis loosening, longitudinal translation and rotation, the measurements for these migration directions are gathered into two groups: clinical and phantom. This overall dataset

is presented in figure 3.2. The standard deviations show that CM-RSA performed significantly better ($p = 0.02$) on the y-rotation than MB-RSA (0.69° vs. 0.96°), while on the clinical dataset, there was no significant difference (0.41 vs 0.35). For the translation along the y-axis, there is a slight increase (from 0.05 mm to 0.06 mm) on the phantom data, while there is a significant increase (from 0.09 mm to 0.14 mm, $p = 0.04$) on the clinical data. EGS-RSA performs similar to MB-RSA (1.02° vs. 0.96°) on the phantom data for the rotation about the longitudinal axis and significantly worse (0.10 mm vs. 0.06 mm, $p < 0.01$) for the translation along the longitudinal axis. On the clinical data, it appeared to perform worse than MB-RSA. Its standard deviations (0.15 mm and 0.79°) were significantly larger ($p = 0.02$ and $p < 0.01$, respectively) than MB-RSA.

For the individual experiments, it can be seen that for the phantom data the mean translation errors were all below 0.1 mm, indicating small systematic errors. For the clinical experiment these values were in general slightly higher, with values below 0.12 mm.

Similarly, the mean rotation errors were mostly below 0.1°, with some exceptions: in the clinical data a mean y-rotation of 0.49° was measured for EGS-RSA applied to the Mallory/Head data. All three methods had a relatively large mean y-rotation on the clinical Stanmore data, with 0.12°, 0.20° and 0.17° for EGS-RSA, MB-RSA and CM-RSA, respectively.

The standard deviations for the translations are in general below 0.2 mm, with the exception of EGS-RSA and CM-RSA applied to the SL-Plus, which show standard deviations for the z-translation of 0.29 mm and 0.49 mm respectively. In the clinical data of the Mallory/Head, EGS-RSA has a standard deviation of 0.30 mm. The standard deviations for the rotations, however, show different results. Although performance on the z-rotations (values between 0.04° and 0.20°) appears to be similar to those of the translations, larger standard deviations can be seen for the x-rotation and y-rotation. In the most extreme case, EGS-RSA has a standard deviation of 2.22° y-rotation when applied to the SL-Plus. Only on the clinical data for the Stanmore, standard deviations comparable to those for translation are obtained, with values of 0.32°, 0.24°, 0.20° for EGS-RSA, MB-RSA, CM-RSA,

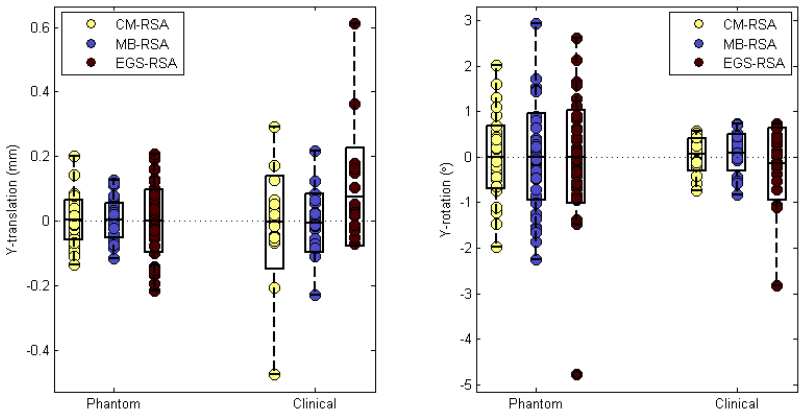


Figure 3.2: The measured Y-translations and Y-rotations for the three methods (EGS-RSA, MB-RSA and CM-RSA). On the left: The Y-translation for the phantom dataset and the clinical dataset. On the right: the Y-rotation for the phantom and clinical dataset. The boxes represent mean standard deviation.

respectively.

For each prosthesis, Levene's test for equality of variances was used to determine if CM-RSA or EGS-RSA showed a significant improvement over MB-RSA. Several significant differences were found. E.g. when considering the y-rotations, EGS-RSA performed significantly better than MB-RSA on both the phantom and clinical data of the Mallory/Head, but at the same time it performed significantly worse on the phantom data for the SL-Plus. CM-RSA performed significantly better on the phantom data for the Stanmore prosthesis.

3.5 Discussion

Overall, the results from figure 3.2 demonstrate that the new CM-RSA method, using a combined stem-head model, can yield more accurate results than the original MB-RSA method with a surface model of the stem only.

When considering the standard deviations for rotation about the y-axis of

EGS-RSA on the phantom data of the Mallory/Head prosthesis, EGS-RSA performs very well, with a factor two improvement for the standard deviation for the rotation about the y-axis (from 1.01° for MB-RSA to 0.51° for EGS-RSA, $p = 0.01$). At the same time, it also may perform better on the phantom data of the Stanmore prosthesis (from 0.94° for MB-RSA to 0.66° for EGS-RSA). The overall worse performance of EGS-RSA with respect to MB-RSA on the phantom data can probably be explained by its poor performance on the SL-Plus prosthesis. On the phantom data of the SL-Plus, EGS-RSA performs a factor two worse ($p < 0.01$) than MB-RSA, with a standard deviation of 2.22° for EGS-RSA, compared to 1.00° for MB-RSA. The SL-Plus stem design has a rectangular cross-section, as opposed to the curved cross-sections of the Mallory/Head and Stanmore designs. For certain stem orientations, this rectangular cross-section can make the estimation of the orientation troublesome. This suggests that the performance of EGS-RSA depends partially on the shape of a particular prosthesis.

In clinical practice the precision is usually worse than the precision in a controlled phantom experiment. For the Mallory/Head and the Stanmore prostheses, this is not visible in the results. The improvement of CM-RSA with respect to MB-RSA as visible in the phantom data, is not clearly present in the clinical data. For the Mallory/Head prosthesis, there is a small non-significant improvement from 0.54° to 0.40° and for the Stanmore prosthesis MB-RSA already achieves good results and CM-RSA yields only a marginal improvement over MB-RSA from 0.24° to 0.204°.

For the Mallory/Head prosthesis, the phantom experiment was performed using a size 9 and a size 12 Mallory/Head prosthesis. During analysis, it appeared that pose estimation on the size 12 prosthesis was much less accurate. Apparently, the MB-RSA and CM-RSA methods have problems with the Mallory/Head size 12 prosthesis, while not having those problems with the other sizes of the Mallory/Head. This can possibly be explained by the fact that the 3D surface model was not reverse-engineered from the actual prosthesis used in the study but from another size-12 prosthesis. This could have resulted in a difference between the 3D surface model of the scanned size 12 Mallory/Head prosthesis and the actual prosthesis used, which in turn could result in relatively large migration errors for the size 12

Mallory/Head prosthesis.

For the discrepancy between phantom and clinical data for the Stanmore prosthesis a simpler explanation can be given. Most Stanmore prostheses in the clinical dataset were standard curved Stanmore prostheses, while the ones from the phantom experiment were straight Stanmore prostheses. The straight Stanmore prosthesis is much more symmetric around its longitudinal axis and will therefore result in a larger standard deviation for the rotation about that axis. As the standard curved Stanmore prosthesis is far less symmetrical and thus less sensitive for measurement error on axial rotation, the addition of the spherical head will not yield a noticeable improvement in the double examinations. This is in line with the initial hypothesis that the addition of the spherical head will yield more precision, because the symmetry along the longitudinal axis is reduced.

The results for the phantom experiment with the SL-Plus prosthesis and the results for the clinical double-examination data for the Mallory/Head prosthesis demonstrate that even with a partial overlap of the head by a metal cup, CM-RSA shows smaller standard deviations (0.84 mm vs. 1.20 mm and 0.40 mm vs. 0.54 mm) for the rotation about the y-axis than MB-RSA. In clinical practice, such a cup will often be present and cause large parts of the spherical head to be occluded.

Kaptein et al. [2006] presented an analysis of the precision of the model-based approach and reported standard deviations for translations ranging from 0.03 mm to 0.21 mm and standard deviations for rotations ranging from 0.04° to 1.76°, with the largest error found for rotation about the y-axis. As can be seen from Table 3.1, the results for MB-RSA presented here are similar.

The results for EGS-RSA are similar to the data presented by Kaptein et al. [Kaptein et al., 2006] with standard deviations for translations ranging from 0.07 mm to 0.14 mm and from 0.10° to 0.61° for rotations. Both results have the largest error for rotation about the y-axis.

The results of CM-RSA can also be compared to the results of marker-based RSA. With standard errors ranging from 0.03 mm to 0.35 mm for

translations [Mjöberg et al., 1986, Kärrholm, 1989, Kärrholm et al., 1994] and from 0.05° to 0.58° for rotations [Kärrholm, 1989, Börlin et al., 2002], the marker-based approach is currently considered the most accurate method for migration measurements.

The results in Table 3.1 show that CM-RSA and EGS-RSA can still have problems with measuring the rotation about the y-axis. Considering the magnitudes of the errors, the small increase in translation error is probably justified by the decrease of the error in the other directions. It can also be seen that there is indeed an increase in accuracy with CM-RSA as opposed to MB-RSA.

3.5.1 Conclusion

It was demonstrated that using a combined stem-head model, with optimisation of the head position during estimation of the pose of a prosthesis, yields more precise migration measurements on the phantom data when compared to a model-based approach with surface models of the stem only. On the same phantom data was demonstrated that elementary geometrical shapes can also be a feasible method for migration measurements on some implant designs.

Overall, CM-RSA appears to be a feasible alternative to MB-RSA. As opposed to CM-RSA and MB-RSA, EGS-RSA eliminates the need for an accurate (reverse engineered) surface model, but it is only applicable in cases where the shape of the implant can be described by elementary geometrical shapes.

Because the precision of the model-based methods - MB-RSA, CM-RSA or EGS-RSA - is shape dependent, it is recommended that before using one of these methods as an alternative for marker-based RSA, an in-vitro validation experiment is carried out to assess the precision of these methods. When the precision of these methods is not sufficient, it is advised to use the marker-based approach, with the disadvantage that markers have to be attached to the prosthesis.

3.5.2 Future Work

One of the reasons for adding the spherical head to the model of the prosthesis was to show that pose estimation using a combined model is feasible. Now that this is demonstrated, the method can be applied to other modular prostheses. E.g. when focusing on knee prostheses, a combined tibial stem-plateau model can be constructed and used to increase the accuracy of RSA.

3.5.3 Acknowledgement

This project was sponsored by European Community project DESSOS IST-2004-27252

CHAPTER 4

DETECTING FEMUR-INSERT COLLISIONS TO IMPROVE PRECISION OF FLUOROSCOPIC KNEE ARTHROPLASTY ANALYSIS

A.H. Prins[1], B.L. Kaptein[1], B.C. Stoel[2], J.H.C. Reiber[2], E.R. Valstar[1,3]

1. *Biomechanics and Imaging Group, Department of Orthopaedics, Leiden University Medical Center, The Netherlands*

2. *Division of Image Processing, Department of Radiology, Leiden University Medical Center, The Netherlands*

3. *Department of Biomechanical Engineering, Faculty of Mechanical, Maritime, and Materials Engineering, Delft University of Technology, The Netherlands*

Journal of biomechanics 2010 43 (4), 694–700

Abstract

Fluoroscopic analysis is an important tool for assessing in vivo kinematics of knee prostheses. Most commonly, a single-plane fluoroscopic setup is used to capture the motion of prostheses during a particular task. Unfortunately, single-plane fluoroscopic analysis is imprecise in the out-of-plane direction. This can result in reconstructing physically impossible poses, in which — for example — the femoral component intersects with the insert, as the normal pose estimation process does not take into account the relation between the components. In the proposed method, the poses of both components are estimated simultaneously, while preventing femur-insert collisions. In a phantom study, the accuracy and precision of the new method in estimating the relative pose of the femoral component were compared to those of the original method. With reverse engineered models, the errors in estimating the out-of-plane position decreased from 2.0 ± 0.7 mm to 0.1 ± 0.1 mm, without effects on the errors in rotations and the in-plane positions. With CAD models, the errors in estimating the out-of-plane position decreased from 5.3 ± 0.7 mm (mean $\pm$ SD) to 0.0 ± 0.4 mm, at the expense of a decreased precision for the other position or orientation parameters. In conclusion, collision detection can prevent reconstructing impossible poses and it improves the position and motion estimation in the out-of-plane direction.

4.1 Introduction

Fluoroscopic analysis is an important tool for assessing in vivo kinematics of knee prostheses. In a typical fluoroscopic setup, the subject performs a certain task (e.g. a step-up, a lunge-motion, etc.), while fluoroscopy is used to capture the internal motion of the prosthesis.

In general, shape matching techniques from the field of computer vision are used to align the shape of a 3D surface model of the implant with its silhouette detected in the image [Banks and Hodge, 1996, Hoff et al., 1998, Zuffi et al., 1999, Kaptein et al., 2003, Tashman and Anderst, 2003, Li et al., 2008]. A virtual contour of a 3D surface model of the prosthesis is calculated and compared with the actual contour of the silhouette, detected in the image. The difference between the virtual and actual contour is minimized using an optimization method.

In a bi-plane setup, high precision motion estimations have been reported [Tashman and Anderst, 2003, You et al., 2001, Kaptein et al., 2003, Li et al., 2008] from 0.06 mm to 0.23 mm for translations and from 0.07 to 1.2 for rotations. In a standard bi-plane setup, however, the freedom of movement for the subject is limited due to the presence of an extra X-ray tube and detector. A single-plane setup allows for sufficient freedom of movement, but it has serious effects on the precision of the shape matching: the out-of-plane precision, with values ranging from 1.4 mm to 4.0 mm, is much worse than the in-plane precision, where absolute precisions in measuring translations from 0.09 mm to 0.46 mm are reported [Banks and Hodge, 1996, Hoff et al., 1998, Mahfouz et al., 2003, Komistek et al., 2003, Kanisawa et al., 2003]. Precisions of the orientation measurement with single-plane fluoroscopy are between 0.35 and 1.3 [Banks and Hodge, 1996, Hoff et al., 1998, Mahfouz et al., 2003, Komistek et al., 2003, Kanisawa et al., 2003], which are comparable to those of bi-plane measurements.

The large error in out-of-plane position estimation may result in an estimated pose, in which the femoral component seemingly intersects with the polyethylene insert, which is physically impossible (see Figure 4.1). It has been proposed [Banks and Hodge, 1996] to ignore the measurement

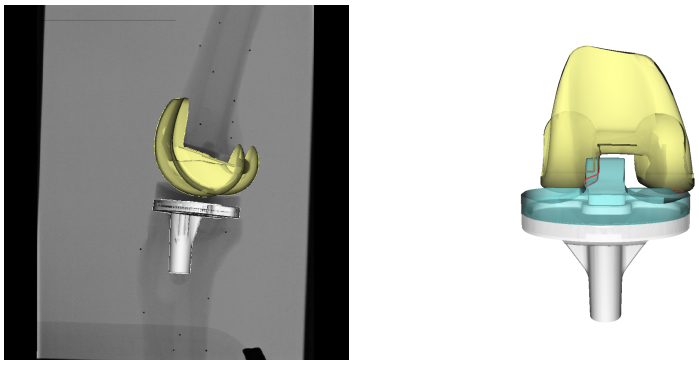


Figure 4.1: Example demonstrating a standard fluoroscopic image with the models overlaid in the poses estimation with normal MBRSA in the left image. The right image shows the displacement with the collision marked.

in the out-of-plane direction and simply use the same z-coordinate as the tibial component. Alternatively, we propose to incorporate collision detection into the pose estimation method. We investigated whether it would be feasible to improve the estimation of out-of-plane positions by preventing femur-insert collisions. It has been demonstrated previously in Roentgen Stereo-photogrammetric Analysis (RSA) how the use of Reverse Engineered (RE) models instead of Computer Aided Design (CAD) models can improve the accuracy of the pose estimation [Kaptein et al., 2003]. To assess the effects of model accuracy on the out-of-plane position we investigated how the normal method and the new method are affected by the use of RE models and CAD models.

The poses were estimated in a single-plane setup by the new method, the normal model-based approach by Kaptein et al. [2003] and the method proposed by Banks and Hodge [1996]. The results of these three methods were compared to those from the much more accurate bi-plane setup, as a gold standard.

4.2 Methods

The new method can be divided into three phases:

1. A combined model is constructed, similar to the models by Prins et al. [Prins et al., 2008];
2. A similarity measure is constructed between the combined model and the images, which takes collisions into account;
3. An optimization method is then used to determine the optimal pose of each component in the combined model.

1. Combined model: A combined surface model [Prins et al., 2008] is constructed from surface models of the femoral and tibial components and the polyethylene insert, with a total of twelve parameters. In this combined model, six parameters are used to control the position and orientation of the femoral component and the remaining six parameters control the position and orientation of the tibial component. Similar to a real implant, the insert in this combined model is placed in a fixed position with respect to the tibial component.

2. Similarity measure: An overall similarity measure is constructed composed of a contour measure and a collision measure. The contour measure indicates the difference between a virtual projection of the model and contours detected in the image, while the collision measure indicates the extent of an intersection between the femoral component and the polyethylene insert.

- Contour measure: Similarly to the normal model-based pose estimation method [Kaptein et al., 2003], contours are detected using a Canny edge detector [Canny, 1986] and the parts corresponding to the femoral and tibial components are selected manually. For each component the average distance is calculated between points on the actual contour detected in the image, and the projected contour of the surface model. This results in two differences, D_f and D_t, for the femoral and tibial component, respectively.
- Collision measure: In addition to the two similarity measures, collision detection is performed using a third-party collision detection library [Terdiman, 2001]. When a collision is present, the intersection between the models for the femoral component and

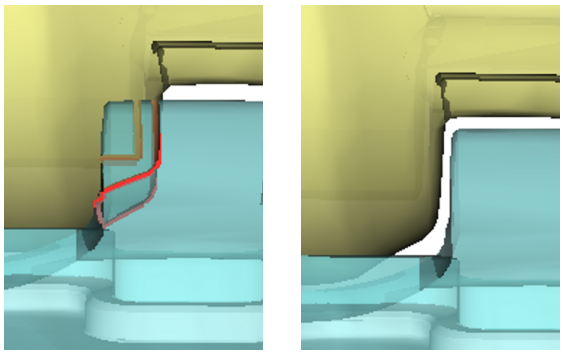


Figure 4.2: Example demonstrating how medial-lateral motion of the femoral component can result in an intersection with the post of the insert. The intersection curve is defined by the locations where the outer surfaces of the femoral component and the insert intersect.

the insert component is composed of a set of volumes. An example of such a situation is presented in Figure 4.2, with a closed curve formed by the surface that the outer surfaces of two components have in common. We use the length of this boundary-curve as an indication of the amount of collision. The sum of the lengths of these curves determines an overall collision measure, D_c, which gives an indication for the extent of the intersection.

- Overall measure: The femoral component is prevented from intersecting with the insert, by using the collision measure as a penalty term. When no intersection is present ($D_c = 0$), there is obviously no penalty And the overall measure is then constructed as $D = D_f + D_t + wD_c$. Pilot experiments showed that $w = 0.5$ was a reasonable value for the weight of the collision in this overall measure.

3. **Optimization:** The optimization is performed in two phases. First the normal model-based approach [Kaptein et al., 2003] is performed on the separate components without collision detection. This ensures that the combined model is initialized to a reasonable position. In a second correction phase, the overall measure D is optimized to refine the estimated pose and to avoid possible collisions. To assess the effects of the different parameters, the following three types of optimizations were performed in this second phase:

- Constrained 2: The relative out-of-plane position is determined by the two z-positions of the femoral and tibial component. The simplest approach is used, which optimizes only those two Z-position parameters, while keeping the other parameters as they were obtained during the first phase.
- Constrained 6: To give the optimizer some additional freedom to correct errors in the relative out-of-plane position and to determine if it will also result in a more consistent in-plane position, all six position-parameters are optimized, whereas the orientation parameters were kept unchanged.
- Constrained 12: To correct for possible small errors in orientation and position and to determine feasibility of a full optimization, optimization on all twelve parameters (controlling pose of both components) is performed. Thus, repeating the entire optimization process based on the initial pose of the first step, but now with additional collision prevention.

4.3 Experimental setup

In a bi-plane flat-panel fluoroscopic setup (Toshiba Infinix-NB: Toshiba, Zoetermeer, The Netherlands), two C-arms were used with the image detectors perpendicular to each other. Using a small calibration box, with a known configuration of embedded markers, the relative X-ray focus positions were calculated [Koning et al., 2007].

The phantom study was performed with a size 3 cruciate-substituting PFC-Sigma prosthesis fixed in sawbones with a 5mm thick insert (DePuy Orthopedics, Warsaw, IN). The tibia-sawbone was fixed with clamps, to prevent the phantom from leaving the field of view. Two motions of the femur were captured with a frame rate of 15 frames per second. In the first motion, the femur moved from full extension to 90° of flexion, followed by a move to approximately 20° abduction, back to 20° adduction and finally back to full extension. In the second motion, the femur started at 30° of flexion, moved to full extension and some internal/external rotation (roughly 20°)

was performed.

To evaluate the effects of model accuracy in a single plane fluoroscopic setup, the analysis was performed with RE and CAD models. The RE models were made from the very same components as were used in the experiment, resulting in the best possible models with respect to accuracy. The following methods were considered:

Reference: As a gold standard measurement, the bi-plane data (from both image detectors) was analyzed with normal model-based pose estimation [Kaptein et al., 2003].

Normal: Normal model-based pose estimation was applied to the single-plane data [Kaptein et al., 2003].

Banks: The results from Normal; followed by a correction to zero of the relative pose of the femoral component [Banks and Hodge, 1996].

Constrained 2: The results from Normal; followed by a correction with optimization of the two z-positions of the components in a combined model with collision detection.

Constrained 6: The results from Normal; followed by a correction with optimization of the six position-parameters of the components in a combined model with collision detection.

Constrained 12: The results from Normal; followed by a correction with optimization of all twelve parameters of the components in a combined model with collision detection.

The precision and accuracy of all five single-plane methods were assessed by determining their difference with the bi-plane measurement Reference: comparing pose of the femoral component, relative to the tibial component. A students T-test and Levene's test for equality of variances were used to check if the various correction methods improved the measurement error with respect to the Normal method.

4.4 Results

A typical result of the constrained pose estimation is presented in Figure 4.3. As measures of accuracy and precision, the mean and standard deviation of the differences in position and in orientation between the methods and the Reference method are presented in Figure 4.4 for the RE models and in Figure 4.5 for the CAD models. Bland-Altman plots [Martin Bland and Altman, 1986] for the relative z-position were used to present the correspondence for the out-of-plane direction (see Figure 4.6). As expected, the Normal method showed a large systematic error (2.0mm for RE models and 5.3mm for CAD models) in the out-of-plane direction.

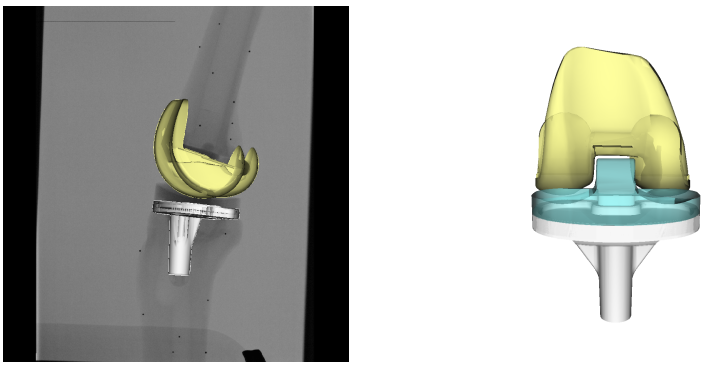


Figure 4.3: Example demonstrating a standard fluoroscopic image with the models overlaid in the poses estimation with constrained MBRSA in the left image. Note that this image is virtually identical to the left image in Figure 4.1. The right image shows the corrected displacement without visible collisions.

With CAD models, the correction of the relative z-position of the femoral component (Banks) reduced this large systematic error significantly (p¡0.001) to 1.1mm, without change in its standard deviation. With RE models, however, it introduced a systematic error of -2.6mm. This error of Banks method is represented as a straight line in Figure 4.6, because it is the direct result of a difference between the actual measurement by the Reference method and the relative z-position of zero imposed by Banks et al.

The three constrained approaches each reduced the systematic out-of-plane error significantly (all three with $p < 0.001$), but to a different extent for RE and CAD models.

With CAD models, the use of only the two z-position parameters (Constrained 2) goes at the expense of a significant increase in the standard deviation (from 0.7 – 2.0 mm, $p = 0.005$). With RE models, the standard deviation for the out-of-plane error improved as well (from 0.7 – 0.3 mm, $p < 0.001$).

The other two constrained methods showed significantly decreased standard deviations for the relative out-of-plane position: Constrained 12 improved to 0.1 mm ($p < 0.001$) with RE models and to 0.5 mm ($p = 0.010$) with CAD models. Constrained 6 improved to 0.1 mm ($p < 0.001$) and to 0.4 mm ($p < 0.001$), for RE and CAD models, respectively.

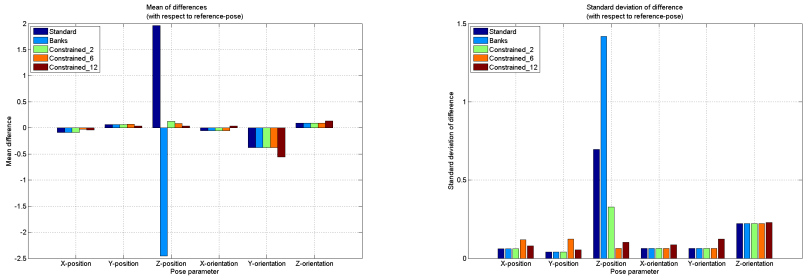


Figure 4.4: Mean (left) and standard-deviations (right) of the differences with RE models from the stereo measurement for the position and orientation of the femoral component relative to the tibial component.

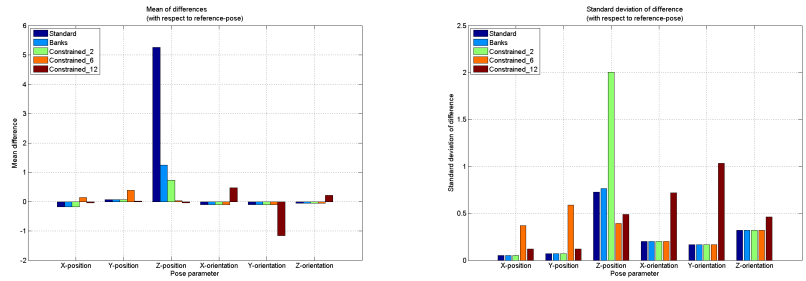


Figure 4.5: Mean (left) and standard-deviations (right) of the differences with CAD models from the stereo measurement for the 24 position and orientation of the femoral component relative to the tibial component.

For the other parameters, there were minor effects with RE models, but large effects with CAD models. There was obviously no difference between the Constrained 2 and Normal methods, neither between the Banks and Normal methods. Comparing Constrained 12 to Constrained 6, there is

a difference with CAD models: it appeared that Constrained 6 achieved lower mean error and standard deviation for the Z-position at the cost of larger errors for the X- and Y-position. The error in X-position changed from -0.1 ± 0.1 mm to 0.1 ± 0.4 mm and for the Y-position it changed from 0.1 ± 0.1 mm to 0.4 ± 0.6 mm. The Constrained 12 method on the other hand showed only significantly smaller mean errors for the X- and Y-positions and slightly larger standard deviations. The largest errors with the Constrained 12 method manifested themselves in the orientations, with significant increases for means and standard deviations of the errors of all orientations. The error in the Y-rotation of $1.2° \pm 1.0°$ in particular was significantly worse ($p < 0.001$ for mean and standard deviations) than the Normal method.

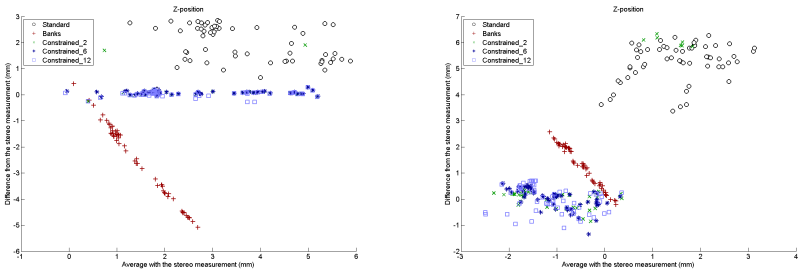


Figure 4.6: Bland-Altman plots for the out-of-plane position of the femoral component, relative to the tibial component. The left plot demonstrates the results with RE models and the right plot those with CAD models. The average with the bi-plane reference measurement is presented on the x-axis with the differences with respect to the reference measurement on the y-axis.

4.5 Discussion

For the out-of-plane Z-position, there is clearly an improvement in accuracy and precision using collision detection in single plane fluoroscopy. With RE models, this is without significant effect on the other parameters, while the results with CAD models demonstrate that improvements are still possible, but at the expense of a decreased precision for the other parameters.

An interesting result is the large systematic error made by the Normal method. With 5.3mm for CAD models and 2.0mm for RE models, such

a large relative displacement of the femoral component of 5.3 mm out-of-plane would imply a femur-insert intersection as demonstrated in Figures 4.1 and 4.2. This large difference is due to the errors in absolute positions of the two components:

The absolute femoral position is biased towards the X-ray focus (3mm with CAD and 0.5mm with RE), while the tibial position is biased towards the image detector (-2mm with CAD and -1.5mm with RE). This opposite shift increases the overall difference out-of-plane between the two components. A systematic error of 5.3 mm with CAD models is much higher than has been reported before [Banks and Hodge, 1996, Mahfouz et al., 2003]. This large error is likely caused by high sensitivity of the estimation of the out-of-plane position to model inaccuracies: when the surface model is slightly smaller than the actual implant the estimate of the out-of-plane position will be further away from the image plane and this can easily be up to a few millimeters.

The large differences in results between CAD and RE models clearly demonstrate that the use of accurate models is very important in a single-plane setup. Note that in clinical practice, the RE models are not made from the very same component as implanted in the patient. So this paper shows the best and the worst case scenario.

Fixing the relative out-of-plane position as applied by Banks improves the accuracy to 1.2 mm for CAD models, but without an improvement of the precision. An additional consequence of fixing the out-of-plane position of the femoral component is that any information on that parameter during a task is lost. We believe this is what occurs with the RE models, where the correction by Banks shows much larger errors (-2.4mm ? 1.4mm).

We can partially explain these worse results of our correction: the correction is based on the assumption that the out-of-plane axis corresponds to the medial-lateral axis of the femorotibial joint and the femoral position is restricted by the surrounding ligaments. Unfortunately, in our experiments this out-of-plane axis was not aligned with the medial-lateral axis (with an angle of 20). Also, by having a contact with the post, the femoral component is in general not precisely centered above the tibia-component and

hence there is a larger out-of-plane difference.

For RE models in general, the three constrained methods show improved out-of-plane accuracies, with a few minor differences for the other parameters. The Constrained 6 method loses some precision in the in-plane directions and the Constrained 12 method loses some precision in the out-of-plane orientations. With either Constrained 6 or Constrained 12, the errors in those other parameters are still very low and therefore the method is suitable for most practical purposes and the lower computation cost of Constrained 6 makes it the method of choice for correcting the out-of-plane position.

With the CAD models, the three constrained methods all show improved out-of-plane accuracies, but each with their own drawback:

- Constrained 2 shows worse precisions for the out-of-plane positions. A possible explanation lies in the fact that the movement in the out-of-plane direction for both components also slightly moves their virtual contours in the X- and Y-direction and the optimization is not capable of correcting for this small movement.

- Constrained 6 shows an improvement of both accuracy and precision in Z-position, but at the expense of a loss in precision in the other relative position-parameters. The cam and post function as an important constraint on the out-of-plane position of both components. In some cases however this means that a change is needed in the in-plane positions to correct for a collision. Combined with discrepancies between the surface models and the actual components, this might explain the change in the in-plane positions. This loss in precision in the in-plane positions need not necessarily be a problem when much larger motions are expected in the in-plane directions. Especially when one is interested in the out-of-plane motion of the femoral component, the decrease in standard deviation for relative out-of-plane from 0.7 mm to 0.4 mm can justify the increase in the other directions.

- Constrained 12 shows an improvement of both accuracy and precision in Z-position comparable to Constrained 6. In theory, it could

be expected to be the most accurate, because all parameters could improve. This is however not the case: larger errors occurred in the out-of-plane orientations. Additionally it is also a much more computationally intensive method, because all twelve parameters must be optimized. Again, the cam and post function as a strong constraint on the out-of-plane positions and the optimizer will try to change the other parameters to satisfy this constraint. Since, changes to the X- and Y-orientation induce only small changes in the projection of the model these provide the optimizer with sufficient freedom to satisfy the collision-constraint.

In previous studies [Tashman and Anderst, 2003, You et al., 2001, Kaptein et al., 2003] precision for in-plane positions in single-plane fluoroscopy was in the range of 0.06mm to 0.23mm. Our results from the Normal method were comparable, but with CAD models, these values could increase to 0.6mm when performing the second-phase optimization with collision prevention.

As far as we are aware, only Hirokawa et al. has used femoral component-insert contact to improve precision of pose estimation [Hirokawa et al., 2008]. By incorporating 3D geometric articulation into the model, they improved the root mean square for the out-of-plane error from 2.6mm, down to 0.8mm, without loss of precision in the other directions. In general all their errors in the relative positions and orientations were below the 1 mm and 1, which are comparable to our results. In their approach femorotibial contact was really necessary, otherwise their improved method would fall back to normal pose estimation. In our approach we forced the components to separate. This has the advantage that the method will work when there is no contact, but at the same time implies that it can end up in a contact-free situation, even though in reality contact was present.

Polyethylene, the material commonly used for the bearing, is sensitive to wear and deformation. This implies that it is difficult to determine the exact thickness of the insert and hence it is difficult to use the articulating surfaces of the condyles for the collision detection, unless one knows the exact geometry and thickness of the articulating surfaces. In many of our studies we perform fluoroscopic analysis from a few months up to a year

postoperatively. In studies with such a follow-up we do not expect significant wear, making the use of collision-detection with the polyethylene insert a useful tool. In those cases where significant wear does occur, this can be detected before the analysis by investigating the distance between the femoral and tibial component.

In this experiment a prosthesis with a cam and post was used. Similar to its function in real life, it can be used to restrict the medial-lateral sliding of the femoral component in pose estimation, which results in an improved estimation of the out-of-plane position. Therefore, the extent of the improvement by our method may depend on the shape of the prosthesis at hand.

Overall, the Constrained 6 method shows the best results. The improvement in the out-of-plane positions is significant, it has a lower computational cost compared to Constrained 12 and with RE models it has no loss of precision in the other parameters. If possible, the use of RE models is recommended, but even with CAD models, the possible loss of precision in the directions other than the out-of-plane direction is small and we are certain that it is sufficiently accurate to detect clinically relevant motions. We conclude therefore that the detection of collisions between an insert and the femoral component can be used effectively to improve the consistency and accuracy of model-based pose estimation for fluoroscopic analysis of total knee prostheses.

4.5.1 Acknowledgement

This project was sponsored by the European Community Project DeSSOS IST-2004-27252.

CHAPTER 5

PERFORMANCE OF OPTIMIZATION IN SINGLE-PLANE FLUOROSCOPIC ANALYSIS FOR TKA

A.H. Prins[1], B.L. Kaptein[1], B.C. Stoel[2], D.J.P. Lahaye, [3], E.R. Valstar[1,4]

1. *Biomechanics and Imaging Group, Department of Orthopaedics, Leiden University Medical Center, The Netherlands*

2. *Division of Image Processing, Department of Radiology, Leiden University Medical Center, The Netherlands*

3. *Department of Applied Mathematics, Faculty Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, The Netherlands*

4. *Department of Biomechanical Engineering, Faculty of Mechanical, Maritime, and Materials Engineering, Delft University of Technology, The Netherlands*

Submitted to Journal of biomechanics 2013

Abstract

Fluoroscopy-derived joint kinematics plays an important role in the evaluation of knee prostheses. Fluoroscopic analysis requires estimation of the 3D prosthesis pose from its 2D silhouette in the fluoroscopic image, by optimizing a dissimilarity measure. Currently, extensive user-interaction is needed, which makes analysis labour-intensive and operator-dependent.

The aim of this study was to review five optimization methods for 3D pose estimation and to assess their performance in finding the correct solution. Two derivative-free optimizers (DHSAnn and IIPM) and three gradient-based optimizers (LevMar, DoNLP2 and IpOpt) were evaluated. For the latter three optimizers two different implementations were evaluated: one with a numerically approximated gradient and one with an analytically derived gradient for computational efficiency.

On phantom data, all methods were able to find the 3D pose within 1 mm and 1° in more than 85% of cases. IpOpt had the highest success-rate: 97%. On clinical data, the success rates were higher than 85% for the in-plane positions, but not for the rotations. IpOpt was the most expensive method and the application of an analytically derived gradients accelerated the gradient-based methods by a factor 3–4 without any differences in success rate.

In conclusion, 85% of the frames can be analyzed automatically in clinical data and only 15% of the frames require manual supervision. The optimal success-rate on phantom data (97% with IpOpt) on phantom data indicates that even less supervision may become feasible.

5.1 Introduction

Single plane fluoroscopy is commonly used to assess the kinematics of knee prostheses and evaluate their design and in-vivo behaviour. To capture the three dimensional (3D) motion of a prosthesis, its position and orientation (pose) are estimated from its silhouette in the individual fluoroscopic frames.

Several methods have been published for estimating the implant pose with reported accuracy of 0.09–0.40 mm for the in-plane position and of 0.35–1.3° for the rotation [Banks and Hodge, 1996, Hoff et al., 1996, Mahfouz et al., 2003, Komistek et al., 2003, Kanisawa et al., 2003, Zuffi et al., 1999, Li et al., 2004, Hermans et al., 2008, Prins et al., 2010]. Although the accuracy is considered sufficient, it is our experience that the analysis is operator-dependent and time-consuming.

Most of the time is spent on the supervised pose estimation where the operator needs to review the results for each frame and restart the estimation process in case of suboptimal solutions. The analysis of hundreds of frames of a single patient can take several hours or days, limiting the reproducibility and applicability of fluoroscopy in larger scale studies.

3D pose estimation from 2D image data can be done based on features, intensities or gradients [Markelj et al., 2012]. Feature-based methods use features extracted from the image as input for the optimization, such as the outer contour of the implant's silhouette. Intensity-based or gradient-based methods perform the estimation directly on the image or gradient data.

In fluoroscopic analysis, a feature-based approach is commonly applied, as the implant features are easily detected in the image. There are two methods of feature-based pose estimation: forward projection and backward projection. In the first method, a projection of a 3D model is made and correspondences between silhouette and projected model points are determined in the image plane. Subsequently, the dissimilarity between silhouette and projection is minimized. The back-projection method determines the correspondences in the image plane too, but then creates projection lines from the silhouette back to the focus and minimizes the dissimilarity between

back-projection lines and the 3D model (see Figure 5.1).

In this study, we applied the back-projection strategy and we defined a nonlinear least squares dissimilarity measure between the back-projection lines and the 3D model. This dissimilarity is then minimized to find the optimal pose of the implant model with respect to the detected silhouette. The classical approach applies Levenberg-Marquardt [Lavallée and Szeliski, 1995, Zuffi et al., 1999, Marquardt, 1963, Levenberg, 1944], but alternative methods have also been proposed [Fregly et al., 2005, Mahfouz et al., 2003].

The accuracy of single-plane fluoroscopic analysis has been assessed only after manual corrections were made. There are no studies indicating the autonomous performance of fluoroscopic analysis , e.g. likelihood of success, convergence rates or computational efficiency.

Therefore the aim of this study was to compare the performance of several optimization methods. We examined derivative-free methods (DHSAnn and IIPM) and gradient-based optimization methods (DoNLP2, IpOpt and Levenberg-Marquardt). For the latter three optimizers two implementations were evaluated: one with a numerically approximated gradient and one with an analytically derived gradient for computational efficiency. The success-rate, dependency on initial pose and the computation timeof each method was investigated in an experiment on phantom and clinical data.

5.2 Methods

To match an implant model to its silhouette in a fluoroscopic image, an accurate 3D surface model and the outer contour of the silhouette is used [Kaptein et al., 2003]. The projection parameters such as focus position and image resolution were determined with Model-based RSA software (Model-Based RSA 3.21, Medis Specials, Leiden, the Netherlands [Kaptein et al., 2003]). The silhouettes were extracted using a Canny edge detector, and the relevant parts on the outer contour were selected by an experienced user.

A 3D surface model is defined by a collection of model points M, and the

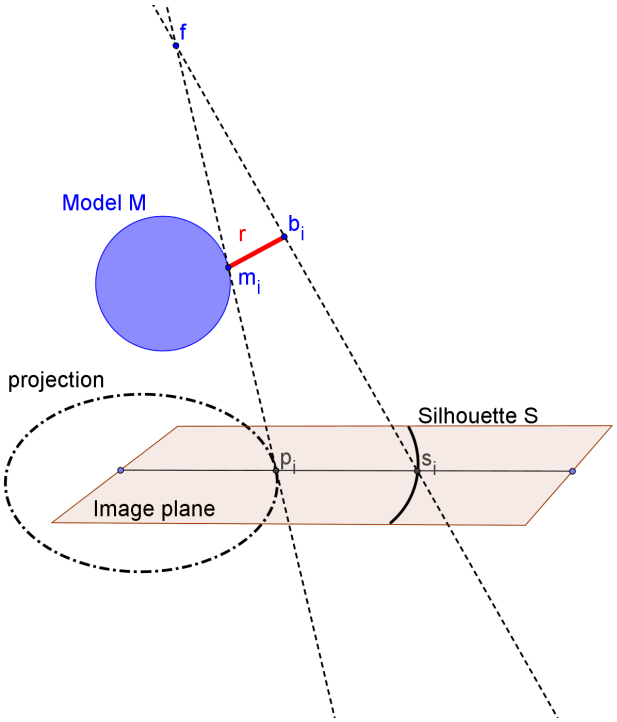


Figure 5.1: Distance measure and projection strategy for pose estimation: ***1:*** *Each model-point m_i is projected onto the image plane.* ***2:*** *The correspondence between projected model-points p_i and silhouette-points s_i is determined in the image-plane.* ***3:*** *Each silhouette points is back-projected to the focus f and the point b_i on the projection line and closest to the corresponding model-point m_i is determined.* ***4:*** *The residual vectors r_i between each projection line and the 3D model points m_i in its candidate pose define the dissimilarity measure.*

2D silhouette by a collection of points S. The estimation of the implant pose ρ' minimizes a dissimilarity measure $\delta(\rho, M, S)$, which indicates how "close" the model M fits the detected silhouette S.

$$\rho' = argmin\ \delta(\rho, M, S) \tag{5.1}$$

A nonlinear least squares dissimilarity measure $\delta(\rho, M, S)$ was defined between model points m_i and their corresponding silhouette points s_i. A generic optimization method can be applied to minimize the dissimilarity measure. The dissimilarity measure and the optimization method are presented in the following two sections.

5.2.1 Dissimilarity Measure

The pose of a 3D implant model is described by six parameters $\rho = (\, x, y, z, \alpha, \beta, \gamma \,)$ which defines a rigid body transformation from a base pose. E.g. applied to each vertex m_i of the 3D model,

$$\phi(\rho, m_i) = R_z(\gamma) \cdot R_x(\alpha) \cdot R_y(\beta) \cdot m_i + (xyz)^T \tag{5.2}$$

where R_x, R_y, R_z are the rotation matrices around the x, y, z axes with the rotation in YXZ-order and $(xyz)^T$ is the translation vector.

The dissimilarity measure $\delta(\rho, M, S)$ from Eq. 5.1 is determined in four steps (see Figure 5.1):

1. Project the implant model in its pose onto the image plane from the focus f as a projected contour P.

2. Determine the correspondences in 2D by finding point pairs (p_i, s_i): the closest point p_i on the projected contour for each detected silhouette point s_i.

3. Define a 3D back-projection line l from each silhouette point s_i to the focus f

   $$b(\lambda) = f + \lambda \cdot (s_i - f) \tag{5.3}$$

   and determine the point b_i on this line closest to the model point m_i, where m_i was the point which resulted in p_i after projection in step 1.

   The point b_i on this projection line closest to the model is defined by the requirement:

   $$(\phi(\rho, m_i) - b_i(\lambda))^T \cdot (s_i - f) = 0 \tag{5.4}$$

   In other words, the vector r_i between b_i and m_i should be perpendicular to the projection line l from f to s_i and λ is calculated as

$$\frac{(\phi(\rho, m_i) - f)^T \cdot (s_i - f)}{(s_i - f) \cdot (s_i - f)} \tag{5.5}$$

4. Define a dissimilarity measure between the projection lines and the corresponding model points as the sum of squared lengths of residual vectors r_i.

$$\delta(\rho) = \sum_{\forall i} \|r_i\|^2 \tag{5.6}$$

The residual vector r_i is computed for each silhouette-point s_i as the shortest vector between the transformed model-point $m_i = \phi(\rho, m_i)$ and the point b_i on the back-projection line l from s_i back towards the focus f:

$$r_i = \phi(\rho, m_i) - b_i \tag{5.7}$$

5.2.2 Gradient

If the derivation of an analytical gradient is feasible, this is often more efficient for gradient-based optimizers. For the aforementioned dissimilarity measure, the gradient is calculated as:

$$\nabla_\rho \delta = 2 \sum_{\forall i} J_\phi \cdot (\phi(\rho, m_i) - b_i) \tag{5.8}$$

Where J is the Jacobian of the rigid body transformation:

$$J_\phi = \left(\frac{\partial \phi}{\partial x} \frac{\partial \phi}{\partial y} \frac{\partial \phi}{\partial z} \frac{\partial \phi}{\partial \alpha} \frac{\partial \phi}{\partial \beta} \frac{\partial \phi}{\partial \gamma} \right) \tag{5.9}$$

The partial derivatives w.r.t. the rotations are given by:

$$\begin{aligned}\frac{\partial\phi}{\partial\alpha} &= R_z(\gamma)\cdot\frac{\partial}{\partial\alpha}R_x(\alpha)\cdot R_y(\beta)\\ \frac{\partial\phi}{\partial\beta} &= R_z(\gamma)\cdot R_x(\alpha)\cdot\frac{\partial}{\partial\beta}R_y(\beta)\\ \frac{\partial\phi}{\partial\gamma} &= \frac{\partial}{\partial\gamma}R_z(\gamma)\cdot R_x(\alpha)\cdot R_y(\beta)\end{aligned} \tag{5.10}$$

whereas the partial derivatives w.r.t. to translations are simply the unit axes:

$$\frac{\partial\phi}{\partial x} = (1\ 0\ 0)^T, \frac{\partial\phi}{\partial y} = (0\ 1\ 0)^T, \frac{\partial\phi}{\partial z} = (0\ 0\ 1)^T \tag{5.11}$$

5.2.3 Optimization methods

Several optimization methods have been published in the context of single-plane fluoroscopic analysis, broadly divided into derivative-free methods and gradient-based methods.

Derivative-free methods require no gradient information, which can be beneficial when the gradient is difficult or expensive to calculate. For a good introduction into recent derivative-free methods, we refer to Conn et al. [2009].

Gradient-based methods require derivatives of the objective function to guide the search towards the minimum. Wright and Nocedal [1999] have provided an overview of general (nonlinear) optimization.

Derivative-free methods

IIPM: The Iterative Inverse Perspective Matching was developed specifically for matching a 3D object to its 2D silhouette [Wunsch and Hirzinger, 1996]. In each iteration the rigid-body alignment between the corresponding points pairs is determined with a landmark transform [Schönemann, 1966]. Our implementation is derived from the implementation in Model-based RSA [Kaptein et al., 2003]

DHSAnn: A combination of Downhill Simplex (Nelder-Mead) and Simulated Annealing [Nelder and Mead, 1965, Kirkpatrick, 1984] examines sev-

eral candidate poses in each iteration with the simplex method. Candidates with an improved cost value are always accepted, but worse candidates may also be selected based on a gradually lowering probability (annealing). The multiple candidates and the occasional acceptance of a "worse" candidate allows the method to jump out of local minima. Our implementation originates from Model-based RSA [Kaptein et al., 2003] and is derived from Numerical Recipes in C [Flannery et al., 1992].

Gradient-based methods

DoNLP2: Sequential Quadratic programming (SQP) is a general method intended specifically for functions which are twice continuously differentiable and can deal with quadratic constraints [Spellucci, 1998]. The method constructs a quadratic model of the dissimilarity measure around the current pose and calculates a new candidate minimizing the quadratic model. We use an implementation applied in Model-based RSA [Kaptein et al., 2003].

In addition to the above three optimizers available in Model-based RSA, we have chosen to investigate the performance of two readily available methods: LevMar and IpOpt.

LevMar: The Levenberg-Marquardt algorithm is a classical iterative approach to solve least-squares problems. In each iteration, it searches for a new candidate with either a Gauss-Newton direction or a gradient descent direction, controlled by a "dampening factor" [Marquardt, 1963, Levenberg, 1944]. Our implementation is based on the "LevMar" software library [Lourakis, Jul. 2004].

IpOpt: IpOpt is a more recent and state-of-the art optimizer, intended for large scale nonlinear optimization of continuous nonlinear problems [Wächter and Biegler, 2006]. It applies an interior-point approach to deal with nonlinear constraints and uses a line-search to compute new candidates.

5.3 Experiments

We performed experiments on high quality bi-plane phantom data and on clinical single-plane data. The phantom data provides high resolution and high contrast image data. In routine clinical practice the image quality can be worse and affect the accuracy of contour detection and pose estimation. The same performance measures were determined on both phantom and clinical data: success-rate, dependency on initial pose, accuracy and computation time.

5.3.1 Phantom data:

Phantom data was collected using a bi-plane flat panel fluoroscopic setup (Super Digital Fluoroscopy (SDF) system, Toshiba Infinix: Toshiba Medical Systems Europe, Zoetermeer, The Netherlands). The experiment was performed with a size 3 cruciate-substituting PFC-Sigma prosthesis fixed in sawbones with a 5 mm thick insert (DePuy Orthopedics, Warsaw, IN). A 3D model of the femoral component was reverse- engineered with an accuracy of 0.05 mm (TNO Industry, Eindhoven, The Netherlands) for pose estimation.

The image detectors were positioned perpendicularly to each other and the sawbones were placed such that one image intensifier had a medial-lateral view, while the other had an anterior-posterior view. The X-ray focus positions were calculated using a calibration box [Kaptein et al., 2011]. A standard model-based pose estimation method (Model-Based RSA 3.21, Medis Specials, Leiden, the Netherlands [Kaptein et al., 2003]) was applied to the images from both image intensifiers and the pose calculated using this approach was used as the reference pose.

5.3.2 Clinical data:

Two fluoroscopic datasets were acquired with 15 fps from a patient performing step-up tasks with a ROCC prosthesis (Biomet, Warsaw, IN) using a single-plane setup (15 fps, Toshiba Infinix, Toshiba, Zoetermeer, the

Netherlands) [Wolterbeek et al., 2012a].

An experienced user detected the implant contours and applied model-based pose estimation on the femoral component with a reverse engineered 3D model [Kaptein et al., 2003, Wolterbeek et al., 2012a]. The resulting poses were considered the reference pose.

5.3.3 Measuring performance

For each frame, the focus position was available and the silhouette and reference pose had been provided by an experienced user. An experienced user can often match an implant model to its silhouette within 5 mm and 10° relative to the optimal pose. Within this region, we generated $N = 10$ random (uniformly distributed) poses for each frame and used these as the initial pose for each estimation method. On the phantom data $282 \times 10 = 2820$ optimizations were performed for each optimizer, and $82 \times 10 = 820$ optimizations on the clinical data.

All optimization methods were configured to halt when the dissimilarity measure drops below 0.0005 mm or when the relative change in dissimilarity measure is smaller than 0.001 mm. A time-limit of five minutes was applied, if this could be configured in the optimizer. Otherwise, each optimization was allowed to run for five minutes in a small pilot experiment and the average number of iterations or function calls was set as the maximum iterations or function calls.

The performance of the optimization methods was investigated in five categories:

Cumulative success-rates are measured for each pose parameter as the percentage of optimizations within a certain acceptable error from the reference pose. At an acceptable error of 0 mm or 0°, the cumulative success-rate starts at 0% and gradually climbs to 100% as the acceptable error increases.

The success-rate of all optimizations combined is presented as "Best-

case". It combines all optimization results ($n = 8$) and picks the optimization result with the lowest dissimilarity. This presents the success-rate feasible in a best-case scenario.

A baseline success-rate is presented as the overall percentage of optimization results within a base-line range of 1 mm in the X-position and Y-position of the reference and within 1° for the rotations. This success-rate will give a practical overall success-rate for each optimization method at clinically relevant accuracies. 1 mm and 1° are commonly reported accuracies of fluoroscopic analysis. The out-of-plane position is commonly inaccurate due to the single-plane nature of the fluoroscopy setup and not considered in the base-line success-rate

Dependency on initial pose is presented as the relation between the dissimilarity measure in the starting pose and in the final, estimated pose. According to the method by van de Kraats et al. [2005], this gives insight into the effect of initial pose on the optimization. Poses contributing to the above base-line success-rate are marked blue, not acceptable poses are marked red.

Accuracy is presented in terms of the error between estimated pose and the reference pose for each pose parameter. For each optimization method, the mean, standard deviation, median, interquartile range and extrema are reported over all frames.

Computation time is presented as the average number of dissimilarity measure evaluations per pose estimation. The average number of seconds to process a frame is reported to give an indication of the processing time.

5.4 Results

5.4.1 Phantom data:

The cumulative success rates are presented (Figure 5.2) and IpOpt showed the best results overall whereas IIPM and LevMar consistently obtained

	IIPM	DHSAnn	LevMar (analytical)	LevMar (numerical)	DoNLP2 (analytical)	DoNLP2 (numerical)	IpOpt (analytical)	IpOpt (numerical)
	Phantom Data							
Success-rate (%)	83.4	93.9	91.3	87.1	94.9	93.5	97.0	97.0
Dissimilarity Evaluations	26	313	60	282	48	231	174	454
CPU-time (s)	0.50	6.11	1.25	12.90	1.19	4.47	2.96	8.09
	Clinical Data							
Success-rate (%)	57.1	84.4	72.3	72.3	79.6	80.9	84.8	84.9
Dissimilarity Evaluations	21	301	57	256	61	270	123	424
CPU-time (s)	0.51	9.80	2.01	19.21	2.15	8.57	2.93	12.70

Table 5.1: Overview of the success rates of each optimization method: the ability of each optimization to result in a pose close to the reference pose: within 1 mm for the X-position and Y-position and within 1° for the rotations. The computation time is presented in terms of mean number of function evaluations per frame and the mean CPU time per frame.

lower success-rates than the other methods. The differences are minimal when comparing the numerical gradient based methods with the corresponding analytical methods, although the use of a numerical gradient seems to have a consistently lower success-rate. The "best-case" success-rate, combining the optimization results of all optimizers, is consistently higher than any other methods. In general all methods, except for IIPM, achieved a high base-line success rate ($> 85\%$). IpOpt has the highest baseline success-rate of 97%.

The relation between the dissimilarity measure in the initial pose and in the final, estimated pose is presented in Figure 5.3. A log-scale was applied on the y-axis which made it easier to distinguish patterns.

Table 5.2 presents the accuracy for each method on phantom data and it shows similar results on the in-plane positions. Larger standard deviations are presented for the out-of-plane Z-position, with IIPM showing the largest value of 6.1 mm. Standard deviations for all rotations range from 0.5° to 2.8°, but the interquartile range is between 0.1° and 0.8°.

The average number of dissimilarity measure evaluations and the average computation time is presented for each optimization method in Table 5.1. The gradient-based methods DoNLP2 and LevMar performed less function evaluations. IpOpt with a numerical approximation showed the largest number of function evaluations. The use of an analytical gradient instead a numerical gradient showed a 3–4 decrease in the number of function evaluations and computation time.

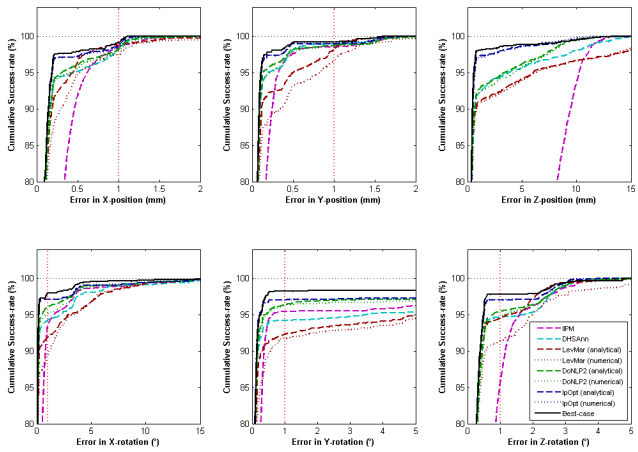


Figure 5.2: Cumulative success-rates (between 80 and 100%) for each pose parameter on experimental data. The bold curve indicates the best-case scenario, where the optimization result was chosen in each frame from the optimization method with the minimal dissimilarity measure. The red vertical lines indicates our baseline success-rate at 1 mm in the X and Y position and at 1° for the rotations.

5.4.2 Clinical data:

Figure 5.4 presents the cumulative success-rates on clinical data. Most methods perform similar in the X-position and X- and Y-rotation, but they show mixed results in the Y-position and Z-rotation. DHSAnn performs better than the other methods on the X-rotation and LevMar performs very well on the Y-position and X-rotation, but poorly on the X-position and Y-rotation. No difference in success-rate was found between a numerical or analytical gradient. The differences between the optimization methods and the best-case success-rates is large on the Y-position, X-rotation and Y-rotation. In the Z-rotation, the LevMar method can outperform the "best-case" scenario.

All methods achieve a high success rate ($> 85\%$) within 1 mm for the in-plane position, but not on the rotations. IpOpt shows the highest baseline success-rate of 85%.

Figure 5.5 presents the relation between the dissimilarity measure at the start of optimization and the dissimilarity measure after optimization.

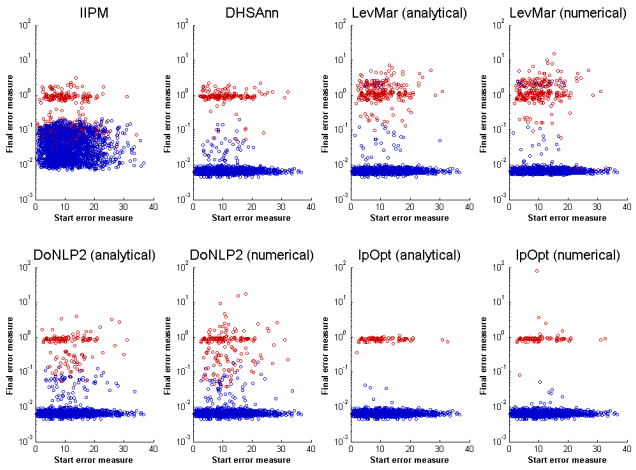


Figure 5.3: Scatter plots with the relation between the dissimilarity measures at the initial pose (x-axis) before the optimization and the dissimilarity measure after optimization (y-axis) for each optimization method on experimental data. The optimization results contributing to the base-line success-rate are marked blue: With an error below 1mm for the X-position and Y-position and below 1° for the X-, Y- and Z-rotation. The remaining optimization results are marked in red.

Large standard deviations for the out-of-plane Z-position (from 4.4 mm to 6.3 mm) can be found (Table 5.3). A large standard deviation of 3° was found for the X-rotation for all optimizers. The interquartile ranges for the X-rotation ranged between 0.5° and 0.8°.

Table 5.1 presents the average number of function evaluations and the average time for optimizing a frame on clinical data. The gradient-based optimization methods show similar computation time when an analytical gradient is applied. The use of an analytical gradient decreased the number of function evaluations by a factor 3–5.

5.5 Discussion

We have investigated the performance of model-based pose estimation for single-plane fluoroscopy using several optimization methods. As far as we know, this is the first study to report success-rates and other performance

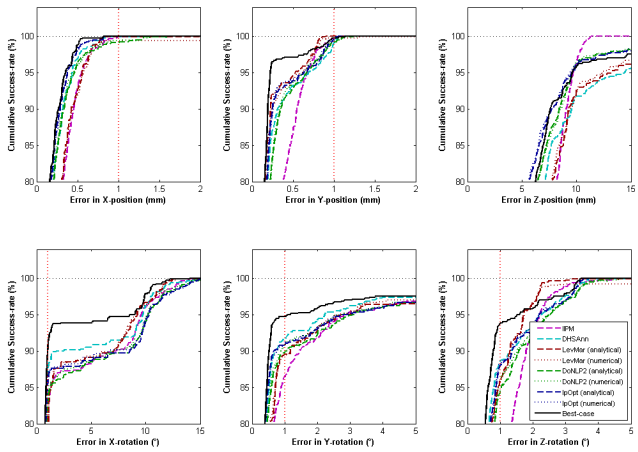


Figure 5.4: Cumulative success-rates (between 80 and 100%) for each pose parameter on clinical data. The bold curve indicates the best-case scenario, where the optimization result was chosen in each frame from the optimization method with the minimal dissimilarity measure. The red vertical lines indicates our baseline succes-rate at 1 mm in the X and Y position and at 1° for the rotations.

data of such methods.

In general, most optimizers obtain correct pose estimates in 85% of cases within 1 mm for the in-plane positions and within 1° for the rotations on phantom data. IpOpt was the most successful method with 97% success on phantom data. However, the success was lower clinically: 85% due to larger errors in the X-rotation. Analytical gradients are a factor 3–4 faster than numerical gradients without loss of accuracy or robustness.

All methods have low success-rates for the out-of-plane Z-position, with the exception of IpOpt on phantom data. The inaccuracy in the out-of-plane position has been reported before and a few methods exist to improve it [Prins et al., 2010, Kanekasu et al., 2004]. Alternatively, the out-of-plane Z-position could be discarded in the analysis [Banks and Hodge, 1996].

Figures 5.3 and 5.5 present horizontal line patterns in the dissimilarity measure before and after pose estimation. The separation between accepted and rejected optimizations in the phantom and clinical data suggests that the final dissimilarity measure after optimization may be a good indicator

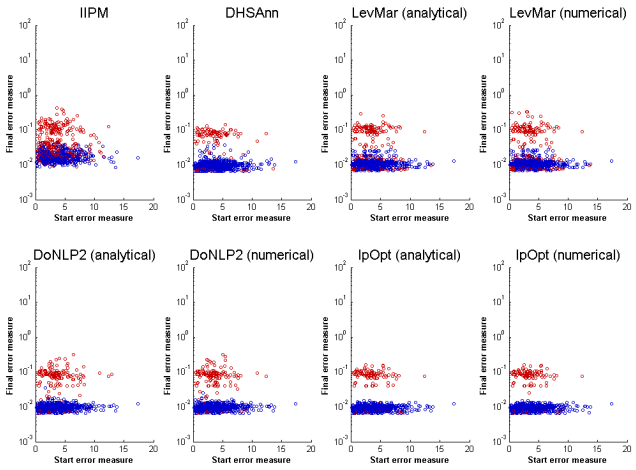


Figure 5.5: Scatter plots with the relation between the dissimilarity measures at the initial pose (x-axis) before the optimization and the dissimilarity measure after optimization (y-axis) for each optimization method on clinical data. The optimization results contributing to the base-line success-rate are marked blue: With an error below 1mm for the X-position and Y-position and below 1° for the X-, Y- and Z-rotation. The remaining optimization results are marked in red.

for success. The vertical spread shown in the plots for DHSAnn and IIPM suggest that these methods have more problems converging to the correct pose.

The clinical data showed that approximately 15% of the optimizations ended up several degrees from the reference pose in the X-rotation, suggesting a second local minimum. The large error in the X-rotation may be caused by the symmetry of the prosthesis component in the sagittal plane [Mahfouz et al., 2003, Hermans et al., 2008].

Median and interquartiles of pose errors are reported, because the errors are not normally distributed. For a fair comparison to the literature the mean and standard deviations were also reported and they are comparable to the literature with errors in the rotations of a few degrees and a few millimeters in the in-plane positions [Banks and Hodge, 1996, Hoff et al., 1998, Zuffi et al., 1999, Mahfouz et al., 2005, Komistek et al., 2003, Kanisawa et al., 2003, Li et al., 2004, Garling et al., 2005].

This study has a few limitations. A uniformly distributed set of random ini-

tial poses was selected. It is unlikely that "experienced" researchers prepare their initial poses randomly, nor is it likely that it is uniformly distributed. Lacking actual user data, it seemed reasonable to start with basic uniform random data.

Results for two prosthesis types were investigated in this study, which may or may not be generalized to other implant types. However, since no data is available on the performance of optimization for other prostheses, our study may provide a good starting point for future comparison of performance data.

The best-case success-rates showed that a clear improvement in the X-rotation and Y-rotation is feasible. A method combining several optimizations may be a feasible, although costly, approach in order to increase the success-rate.

An alternative approach may be automatic detection of unacceptable results, perhaps based on the dissimilarity measure, after which a restart of optimization can be triggered or a different method may be attempted.

Conclusion

This study presents an overview of the robustness, accuracy and computational efficiency of several 3D pose optimization methods for single-plane fluoroscopy. IpOpt performs best and a speedup by a factor 3–4 was shown from an analytical gradient. 85% of the frames can be analyzed automatically in clinical data and only 15% of the frames require manual supervision. The optimal success-rate on phantom data (97% with IpOpt) on phantom data indicates that even less supervision may become feasible.

	IIPM					
	X-position (mm)	Y-position (mm)	Z-position (mm)	X-rotation (°)	Y-rotation (°)	Z-rotation (°)
mean :	-0.068	0.040	-0.006	-0.203	0.375	-0.207
std :	0.305	0.228	6.098	1.474	1.975	0.793
med :	-0.028	0.010	-0.161	0.023	-0.009	-0.093
iqr :	0.335	0.171	10.364	0.370	0.266	0.820
min :	-2.085	-0.337	-12.681	-20.655	-0.777	-4.800
max :	0.704	1.867	13.678	3.638	15.062	1.575
	DHSAnn					
	X-position (mm)	Y-position (mm)	Z-position (mm)	X-rotation (°)	Y-rotation (°)	Z-rotation (°)
mean :	-0.059	0.019	-0.164	-0.179	0.604	-0.167
std :	0.218	0.184	2.165	1.635	2.836	0.613
med :	-0.019	-0.011	0.104	0.006	-0.021	-0.064
iqr :	0.092	0.063	0.395	0.083	0.082	0.307
min :	-1.428	-0.330	-16.718	-19.440	-0.671	-4.157
max :	2.012	1.757	13.422	7.761	22.349	3.816
	LevMar (analytical)					
	X-position (mm)	Y-position (mm)	Z-position (mm)	X-rotation (°)	Y-rotation (°)	Z-rotation (°)
mean :	-0.027	0.038	-0.416	-0.016	0.446	-0.136
std :	0.243	0.230	3.381	1.752	2.192	0.578
med :	-0.013	-0.009	0.070	0.025	-0.022	-0.057
iqr :	0.095	0.066	0.397	0.087	0.089	0.316
min :	-2.897	-1.842	-32.711	-19.348	-7.063	-5.434
max :	3.754	1.903	22.764	10.300	15.120	6.457
	LevMar (numerical)					
	X-position (mm)	Y-position (mm)	Z-position (mm)	X-rotation (°)	Y-rotation (°)	Z-rotation (°)
mean :	-0.020	0.023	-0.489	-0.013	0.450	-0.085
std :	0.317	0.316	3.628	1.760	2.248	0.968
med :	-0.016	-0.008	0.085	0.021	-0.026	-0.065
iqr :	0.095	0.067	0.404	0.090	0.096	0.330
min :	-2.606	-2.142	-43.898	-19.368	-7.846	-5.535
max :	5.550	3.065	21.568	10.537	15.122	8.491
	DoNLP2 (analytical)					
	X-position (mm)	Y-position (mm)	Z-position (mm)	X-rotation (°)	Y-rotation (°)	Z-rotation (°)
mean :	-0.057	0.013	0.038	-0.158	0.290	-0.147
std :	0.205	0.205	1.730	1.324	1.849	0.589
med :	-0.017	-0.011	0.101	0.011	-0.023	-0.061
iqr :	0.091	0.066	0.391	0.084	0.079	0.316
min :	-2.338	-2.537	-11.750	-19.770	-1.598	-5.597
max :	1.282	2.426	16.009	5.446	15.421	2.686
	DoNLP2 (numerical)					
	X-position (mm)	Y-position (mm)	Z-position (mm)	X-rotation (°)	Y-rotation (°)	Z-rotation (°)
mean :	-0.059	0.015	0.034	-0.151	0.302	-0.164
std :	0.240	0.229	1.795	1.348	1.870	0.612
med :	-0.021	-0.011	0.116	0.007	-0.023	-0.068
iqr :	0.094	0.066	0.397	0.085	0.079	0.323
min :	-4.534	-3.057	-11.497	-19.678	-7.030	-5.870
max :	3.568	3.066	19.663	8.476	15.290	3.586
	IpOpt (analytical)					
	X-position (mm)	Y-position (mm)	Z-position (mm)	X-rotation (°)	Y-rotation (°)	Z-rotation (°)
mean :	-0.050	0.007	-0.011	-0.158	0.274	-0.122
std :	0.164	0.169	1.227	1.307	1.831	0.493
med :	-0.017	-0.012	0.105	0.011	-0.023	-0.059
iqr :	0.088	0.063	0.354	0.078	0.076	0.294
min :	-1.257	-0.182	-12.129	-19.678	-1.153	-4.630
max :	0.316	1.864	16.204	0.395	15.428	0.602
	IpOpt (numerical)					
	X-position (mm)	Y-position (mm)	Z-position (mm)	X-rotation (°)	Y-rotation (°)	Z-rotation (°)
mean :	-0.052	0.011	-0.003	-0.154	0.286	-0.121
std :	0.176	0.275	1.141	1.265	1.863	0.500
med :	-0.019	-0.012	0.115	0.008	-0.023	-0.064
iqr :	0.090	0.061	0.372	0.079	0.076	0.293
min :	-1.675	-0.182	-11.192	-19.678	-0.505	-4.584
max :	2.238	11.417	19.100	0.539	15.272	6.505

Table 5.2: Accuracy on phantom data: mean, standard deviation, median, interquartile range, minima and maxima of the error in pose w.r.t. the reference pose.

	IIPM					
	X-position (mm)	Y-position (mm)	Z-position (mm)	X-rotation (°)	Y-rotation (°)	Z-rotation (°)
mean :	0.017	0.145	0.037	-0.994	-0.353	-0.075
std :	0.271	0.278	6.029	3.282	1.635	1.100
med :	0.021	0.103	0.158	0.144	0.039	-0.088
iqr :	0.325	0.331	10.732	0.851	0.589	1.376
min :	-0.814	-0.873	-11.292	-19.515	-15.649	-3.120
max :	1.155	1.154	11.239	1.534	2.117	4.654
	DHSAnn					
	X-position (mm)	Y-position (mm)	Z-position (mm)	X-rotation (°)	Y-rotation (°)	Z-rotation (°)
mean :	0.013	0.091	-1.537	-0.712	-0.225	-0.092
std :	0.197	0.242	6.225	3.059	1.478	0.875
med :	-0.002	0.036	-2.121	0.203	0.081	-0.052
iqr :	0.157	0.175	5.701	0.580	0.337	0.581
min :	-0.634	-0.462	-25.249	-14.939	-10.106	-3.005
max :	1.584	1.139	23.349	1.358	1.125	4.990
	LevMar (analytical)					
	X-position (mm)	Y-position (mm)	Z-position (mm)	X-rotation (°)	Y-rotation (°)	Z-rotation (°)
mean :	-0.010	0.077	-2.408	-0.706	-0.255	-0.034
std :	0.265	0.191	5.911	3.137	1.739	0.777
med :	-0.021	0.039	-2.917	0.298	0.161	0.018
iqr :	0.238	0.180	5.915	0.777	0.404	0.638
min :	-0.812	-0.337	-21.263	-19.823	-15.976	-2.258
max :	2.789	0.899	26.021	1.598	1.993	7.644
	LevMar (numerical)					
	X-position (mm)	Y-position (mm)	Z-position (mm)	X-rotation (°)	Y-rotation (°)	Z-rotation (°)
mean :	-0.000	0.086	-2.105	-0.697	-0.228	-0.027
std :	0.312	0.224	5.992	3.135	1.676	0.931
med :	-0.020	0.038	-2.556	0.290	0.142	0.016
iqr :	0.233	0.175	6.075	0.790	0.399	0.630
min :	-0.810	-0.344	-20.570	-19.775	-16.386	-7.959
max :	2.826	2.913	30.012	2.809	2.712	7.653
	DoNLP2 (analytical)					
	X-position (mm)	Y-position (mm)	Z-position (mm)	X-rotation (°)	Y-rotation (°)	Z-rotation (°)
mean :	0.006	0.091	-1.784	-1.062	-0.326	-0.094
std :	0.225	0.244	4.838	3.487	1.690	0.970
med :	0.001	0.039	-1.761	0.139	0.069	-0.065
iqr :	0.160	0.189	5.278	0.650	0.337	0.579
min :	-1.599	-0.657	-19.856	-19.743	-16.405	-3.858
max :	1.305	1.236	19.272	3.623	1.563	4.479
	DoNLP2 (numerical)					
	X-position (mm)	Y-position (mm)	Z-position (mm)	X-rotation (°)	Y-rotation (°)	Z-rotation (°)
mean :	-0.000	0.079	-1.840	-0.999	-0.282	-0.102
std :	0.217	0.240	4.858	3.406	1.679	0.962
med :	-0.000	0.040	-1.736	0.163	0.077	-0.057
iqr :	0.157	0.192	5.853	0.619	0.326	0.612
min :	-1.503	-0.681	-19.954	-19.715	-16.493	-4.740
max :	1.274	1.313	19.224	1.733	1.706	4.612
	IpOpt (analytical)					
	X-position (mm)	Y-position (mm)	Z-position (mm)	X-rotation (°)	Y-rotation (°)	Z-rotation (°)
mean :	0.008	0.082	-1.976	-0.979	-0.297	-0.103
std :	0.169	0.221	4.457	3.491	1.677	0.875
med :	-0.004	0.034	-1.893	0.193	0.082	-0.065
iqr :	0.138	0.161	4.511	0.549	0.279	0.536
min :	-0.612	-0.295	-19.903	-19.715	-16.574	-2.891
max :	0.814	1.052	19.223	1.340	1.600	3.714
	IpOpt (numerical)					
	X-position (mm)	Y-position (mm)	Z-position (mm)	X-rotation (°)	Y-rotation (°)	Z-rotation (°)
mean :	0.011	0.078	-2.028	-0.941	-0.280	-0.094
std :	0.170	0.217	4.377	3.448	1.672	0.869
med :	-0.002	0.035	-1.916	0.194	0.089	-0.063
iqr :	0.138	0.168	4.654	0.535	0.294	0.544
min :	-0.584	-0.292	-19.903	-19.715	-16.594	-2.830
max :	1.128	1.048	19.220	1.338	1.570	3.737

Table 5.3: Accuracy on patient data: mean, standard deviation, median, interquartile range, minima and maxima of the error in pose w.r.t. the reference pose.

CHAPTER 6

DETECTING CONDYLAR CONTACTLOSS USING SINGLE-PLANE FLUOROSCOPY: A COMPARISON WITH *IN VIVO* FORCEDATA AND *IN VITRO* BI-PLANE DATA

A.H. Prins[1], B.L. Kaptein[1], S.A. Banks[3], B.C. Stoel[2], R.G.H.H. Nelissen[1], E.R. Valstar[1,4]

1. *Biomechanics and Imaging Group, Department of Orthopaedics, Leiden University Medical Center, The Netherlands*

2. *Division of Image Processing, Department of Radiology, Leiden University Medical Center, The Netherlands*

3. *Department of Mechanical & Aerospace Engineering, University of Florida, Gainesville, Florida*

4. *Department of Biomechanical Engineering, Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, The Netherlands*

Journal of biomechanics 2014 47 (7), 1682–1688

Abstract

Knee contact mechanics play an important role in knee implant failure and wear mechanics. Femoral condylar contact loss in total knee arthroplasty has been reported in some studies and it is considered to potentially induce excessive wear of the polyethylene insert. Measuring in vivo forces applied to the tibial plateau with an instrumented prosthesis is a possible approach to assess contact loss in vivo, but this approach is not very practical. Alternatively, single-plane fluoroscopy and pose estimation can be used to derive the relative pose of the femoral component with respect to the tibial plateau and estimate the distance from the medial and lateral parts of the femoral component towards the insert. Two measures are reported in the literature: lift-off is commonly defined as the difference in distance between the medial and lateral condyles of the femoral component with respect to the tibial plateau; separation is determined by the closest distance of each condyle towards the polyethylene insert instead of the tibia plateau. In this validation study, lift-off and separation as measured with single-plane fluoroscopy are compared to in vivo contact forces measured with an instrumented knee implant. In a phantom study, lift-off and separation were compared to measurements with a high quality bi-plane measurement. The results of the in vivo contact-force experiment demonstrate a large discrepancy between single-plane fluoroscopy and the in vivo force data: single-plane fluoroscopy measured up to 5.1mm of lift-off or separation, whereas the force data never showed actual loss of contact. The phantom study demonstrated that the single-plane setup could introduce an overestimation of $0.22\text{mm} \pm 0.36\text{mm}$. Correcting the out-of-plane position resulted in an underestimation of medial separation by $-0.20\text{mm} \pm 0.29\text{mm}$. In conclusion, there is a discrepancy between the in vivo force data and single-plane fluoroscopic measurements. Therefore contact loss may not always be determined reliably by single plane fluoroscopy analysis.

6.1 Introduction

The study of contact mechanics plays an important role in investigating polyethylene wear and implant failure in total knee arthroplasty. Contact loss between the femoral condyles and the polyethylene inlay may result in excessive loading on the side that retains contact and in high impact forces on the side that loses and regains contact. Consequently, contact loss may be related to excessive wear of the polyethylene insert in the knee implant [Andriacchi, 1994, Nilsson and Kärrholm, 1993, Dennis et al., 2001].

Some researchers have used pressure films during surgery to assess contact profiles [Sharma et al., 2007], but this cannot be used to assess post-operative contact. Alternatively, a few instrumented knee implants have been used to assess the in vivo contact force during dynamic activities [Heinlein et al., 2007, Zhao et al., 2007]. However, both methods cannot be applied on a large scale and are not applicable in a general clinical setting.

Single-plane fluoroscopy with 3D pose estimation techniques can be used to determine the relative position and orientation (pose) of the femoral component with respect to the tibial plateau [Dennis et al., 1996, Banks and Hodge, 1996, Komistek et al., 2003, Kaptein et al., 2003]. Pose estimation from single-plane fluoroscopic data has an accuracy ranging from 0.09mm to 0.40mm for the two in-plane positions and from 0.35° to 1.3° for all three orientations [Banks and Hodge, 1996, Hoff et al., 1998, Mahfouz et al., 2003, Komistek et al., 2003, Kanisawa et al., 2003, Garling et al., 2005]. With only a single X-ray focus, the accuracy in the out-of-plane position (medial-lateral), which can be up to 5 mm, is low compared to other directions [Prins et al., 2010].

With relative poses for the femoral and tibial components, the distance between the femoral condyles and the tibial plateau can be measured and two measures for contact loss have been presented in the literature: lift-off and separation.

For determining lift-off, the distances of the lowest point on each condyle with respect to the tibial plateau is calculated (Figure 6.1). Lift-off is then

determined as the difference between these two distances. [Stiehl et al., 1999, Dennis et al., 2001, Insall et al., 2002]. It is a fairly straightforward measure, easy to calculate, and relies only on the relative orientation of the femoral component with respect to the tibial plateau. It neglects, however, the curved surface of the insert and, to distinguish contact loss from measurement error, a threshold of $0.5 - 1.0$mm [Stiehl et al., 1999, Dennis et al., 2001] must be applied.

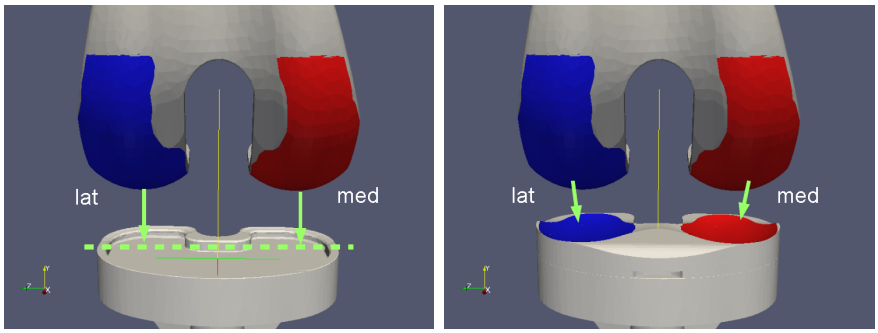


Figure 6.1: Lift-off (left figure) is calculated as the difference in distance between the lowest points on the medial and lateral condyles of the femoral component and the plane through the tibia-plateau. Separation (right figure) is calculated as the closest distance between a point on a condyle and a point on the corresponding part of the polyethylene insert.

Separation is defined as the closest distance (Figure 6.1) between the femoral condyles and the polyethylene insert [Kanekasu et al., 2004]. In theory, this is a more accurate measure, because it considers the shape of the insert and femoral condyles. However, the insert is not visible on fluoroscopy and a model must be used, and this is more sensitive to errors in the pose of the femoral component and the tibia plateau (i.e. the tibial component with its insert). We suspect that the threshold of $0.5 - 1.0$mm may be too low to accurately distinguish loss of femorotibial contact from measurement error and that the large error in out-of-plane position may introduce additional error, especially when considering the curvature of the insert.

This study aims to assess the feasibility of using lift-off and separation as a surrogate measure for contact loss, when derived from single-plane fluoroscopy. In vivo force data from a patient with an instrumented knee prosthesis were collected and the medial and lateral forces on the tibial plateau were compared to lift-off and separation as measured with single-

plane fluoroscopic analysis. In order to explore the results from this in vivo experiment further, the sensitivity of lift-off and separation to pose estimation errors was studied: In an additional bi-plane phantom experiment single-plane fluoroscopy is compared to bi-plane fluoroscopy evaluating the differences between lift-off and separation with high quality image data.

6.2 Methods

In each frame in a fluoroscopic examination, contact loss is detected by estimating the 3D pose of each component and subsequently calculating lift-off and separation.

6.2.1 Pose estimation

The pose of the prosthetic components can be estimated using various methods which have different accuracies, especially in the out-of-plane position [Prins et al., 2010]. As this affects the accuracy of separation measurements, we applied two pose estimation methods for fluoroscopy:

Standard: A model-based pose estimation method minimizing the difference of the virtual projected contour of the implant model with the detected contour of the implant in the fluoroscopic image [Kaptein et al., 2003, Prins et al., 2010].

Corrected: The same method as the *Standard* method was applied, but to reduce errors in the relative out-of-plane position, the femoral component was translated along the out-of-plane axis and centered above the tibial component [Banks and Hodge, 1996, Prins et al., 2010].

With both single-plane methods, the poses of the tibial component and the femoral component were estimated for each image frame. The poses of the insert were fixed with respect to the tibial component.

For the validation of the above methods in the phantom experiment, gold standard bi-plane data were obtained by model-based pose estimation (Model-based RSA 3.21, Medis specials, Leiden, the Netherlands [Kaptein et al., 2003]), providing a very accurate pose measurement without a large error in the out-of-plane position. To ensure the highest possible accuracy, a reverse engineered (laser scan) model was used for the phantom experiment. The model was reduced to 5000 triangles to reduce computation times. A CAD model was used in the in vivo experiment and reduced to 5000 triangles as well [Kaptein et al., 2003].

6.2.2 Contact loss detection

To measure contact loss, two different measures are available with different nomenclature. In this paper we considered lift-off as defined by Stiehl et al. [1999] and separation as expressed in Kanekasu et al. [2004].

Lift-off: For lift-off, the points on the medial and lateral condyles closest to the plane through the tibial plateau were determined (Figure 6.1). This results in a distance hmed and hlat for each condyle. Their difference is defined as:

$$\begin{aligned} \text{medial lift-off} = h_{med} - h_{lat} &\quad \text{, if } h_{med} > h_{lat} \\ 0 &\quad \text{, if } h_{med} \leq h_{lat} \\ \text{lateral lift-off} = h_{lat} - h_{med} &\quad \text{, if } h_{lat} > h_{med} \\ 0 &\quad \text{, if } h_{lat} \leq h_{med} \end{aligned}$$

Note that this implies that medial and lateral lift-off are mutually exclusive and that it is influenced only by the relative orientation of the femoral component with respect to the tibial component and not on the insert.

Similar to Stiehl et al. [1999], we considered lift-off above 1.0mm to represent actual contact loss, and values below 1.0mm as measurement errors.

Separation: Separation takes the insert geometry into account with the insert position fixed to the tibia plateau. The distance is calculated for the medial and lateral part as follows (Figure 6.1):

1. For each point p on the insert, the closest distance $d(p)$ towards a face of the corresponding part on the femoral component was calculated.
2. For each point q on the femoral component, the closest distance $d(q)$ towards the corresponding part on the insert was calculated.

Separation is determined as the minimal distance

$$\text{Separation} = min\left(min\left(d(p)\right), min\left(d(q)\right)\right)$$

Note that separation is not mutually exclusive, since both medial and lateral separation can occur simultaneously. We also used a threshold of 1.0mm to distinguish actual contact loss from measurement errors.

6.3 In vivo experiment

The in vivo data used in this study were collected in a previous study by Zhao et al. [2007] and consist of in vivo force data from a patient with a custom instrumented knee implant and in vivo fluoroscopic data Zhao et al. [2007]. The instrumented knee implant has four uniaxial force sensors at known locations [DLima et al., 2005]. We calculated the medial force as the sum of the forces applied to the two medial sensors, and lateral force as sum of the forces on the lateral sensors. In order to determine the predominant location of the contact forces, the relative medial force was calculated as the percentage of the total force that was medial. We compared the medial and lateral contact force data to lift-off and separation from fluoroscopic data.

Simultaneously with collecting the force data, lateral fluoroscopic images were acquired (Precise Optics P1808 C-arm, 23cm image intensifier, continuous beam of 75kVp and 1mA using an electronically shuttered video camera with $1 - 2$ms exposures). Images were recorded to digital video

tape, transferred to a computer and corrected for geometric distortion using bilinear interpolation [Banks and Hodge, 1996].

During the force and fluoroscopic data acquisition, the patient performed a variety of tasks. Four fast and four slow gait cycles on a treadmill were collected. The full dataset with force data was recorded. The fluoroscopic data were recorded at 30 frames per second (fps) and between 15 and 25 fluoroscopic frames surrounding heelstrike were manually selected based on flexion angle for each gait cycle. Similarly, $10 - 15$ frames were selected around toe-off in each gait-cycle.

Three dynamic step-up activities were collected with approximately 30 frames for each step-up (10fps). Stair stepping was performed with the subjects foot on a 20cm riser with the toes pointed directly forward. Images were recorded as the subject stepped up directly into full weightbearing extension on the replaced knee, without swinging through the opposite leg, and then immediately reversed direction and lowered themselves to rest upon the contralateral leg. The subject was offered hand support for balance but could not lift with their arms.

Two static activities, kneel and lunge, were collected, with approximately 20 frames each (10fps): Kneeling was performed with the implanted knee placed on a padded chair at approximately 90° flexion, while the extended contralateral limb supported most of the body weight. The subject was asked to bend from 90° flexion to maximum comfortable flexion while lateral fluoroscopic images were recorded. Lunging was performed with the subjects foot placed on a 20cm riser. The subject was asked to slowly bend to maximum comfortable knee flexion, in an exaggerated shoe-tying posture, while images were recorded. Their motions were not constrained and the subject was allowed to lift their heel if that permitted a greater range of flexion. An investigator offered to hold the subject's hands or forearms as a safety measure to prevent a fall.

Unfortunately, it was not possible to synchronize in time the data collection of the contact force from the instrumented knee and the collection with single-plane fluoroscopy. This makes a fine-grained comparison difficult between fluoroscopy and force. Instead, we report over complete trials

the mean and minimum medial and lateral force. If there has been loss of contact, this should show as a low minimum force and as high lift-off and separation in the fluoroscopic data.

The accuracy of the force sensors has been reported to be in the range of $0.3 - 3.2\%$ with load experiments at $178 - 712$N [Kaufman et al., 1996]. A worst-case error of 3.2% at the highest tested load of 712N, would imply a worst-case error bound of 23N. At the start of this experiment we decided on a safe threshold of 50N to detect actual contact and to prevent detecting false-positive contact loss.

In summary, a low medial force (below 50N) suggests medial contact loss and similarly for the lateral side, less than 50N indicates lateral contact loss. This is compared to the maximum medial or lateral, lift-off and separation from single-plane fluoroscopy: High lift-off or separation suggesting contact loss. An equal medial and lateral force from the instrumented knee implies that no contact loss has occurred and thus no large lift-off or separation values should be measured from the fluoroscopic data.

6.4 Phantom experiments

We performed two experiments with highly accurate data from a bi-plane setup, to assess the effect of errors in the out-of-plane position. The first experiment compares *bi-plane lift-off* to *bi-plane separation* and measures their difference, independent of the error in the out-of-plane position. The second experiment evaluates the effect of out-of-plane position error from single-plane fluoroscopy on lift-off and separation.

The experimental setup consisted of a bi-plane flat panel fluoroscopic system (Super Digital Fluoroscopy (SDF) system, Toshiba Infinix: Toshiba, Zoetermeer, The Netherlands). The image sensors were positioned perpendicular to each other and the X-ray focus positions relative to the image plane were calculated using a calibration box [Koning et al., 2007, Kaptein et al., 2011].

The phantom experiment was performed with a cruciate-substituting medium

size prosthesis fixed in sawbones with a 5mm thick tibial insert (PFC-Sigma CS, DePuy Orthopaedics, Warsaw, IN). The tibia-sawbone was fixed with clamps, to prevent the phantom from leaving the field of view.

Two motions of the femur were captured with 15fps: In the first motion, the femur moved from full extension to 90° of flexion, then to approximately 20° abduction, back to 20° adduction and finally back to full extension. In the second motion, the femur started at 30° of flexion, moved to full extension and some internal/external rotation (roughly 20°) was performed. In this experiment there was actual contact loss at the medial condyle through parts of both runs, while we tried to keep the lateral condyle in contact with the insert. 282 frames were collected and used for both experiments, 146 frames in the first and 136 in the second run.

6.4.1 Phantom experiment 1: lift-off vs. separation

The first experiment was performed comparing *bi-plane lift-off* directly to *bi-plane separation*. Paired students t-test was used to test for significant differences. We assumed that the accuracy of pose estimation is sufficiently high with bi-plane fluoroscopy that the differences between the measurements can be attributed to the differences between the methods.

6.4.2 Phantom experiment 2: effects of measurement accuracy

The second experiment investigated the effect of single-plane measurement accuracy. The *bi-plane separation* is derived from femoral and tibial poses from high quality bi-plane data and it takes the insert-shape correctly into account when measuring contact loss. This makes *bi-plane separation* our *Bi-plane separation reference* measurement. Similar to the *in vivo* experiment, we calculated *lift-off* and *separation* each with single plane pose estimation methods: *Standard* and *Corrected*. The differences with respect to *Bi-plane separation reference* were calculated and a Students t-test was used to test for significant differences.

6.5 Results

In Vivo Experiment

	Medial (N)			Lateral (N)			Relative Medial (%)		
	Mean	Std	Min	Mean	Std	Min	Mean	Std	Min
Heelstrike	452	323	**39**[1]	388	174	87	48	13	12
Toe-off	429	329	**11**[2]	376	173	76	46	14	3
Step-up	813	258	234	705	239	206	54	7	28
Kneel	116	17	86	88	11	68	57	3	47
Lunge	607	109	474	433	65	356	58	1	56

[1]Medial contact loss in heelstrike, consistent with lift-off/separation table 6.2
[2]Medial contact loss in toe-off, disagreeing with the force data in table 6.2

Table 6.1: Mean, standard deviation and minima for medial and lateral force and for the relative medial force for five datasets. Values in bold are below the predefined threshold of 50N.

	Medial (mm)			Lateral (mm) (%)		
	Lift-off	Separation (*Standard*)	Seperation (*Corrected*)	Lift-off	Separation (*Standard*)	Seperation (*Corrected*)
Heelstrike[1]	**5.11**[1]	**2.24**[1]	**2.60**[1]	**5.06**	**4.53**	**4.04**
Toe-off	0.59	0.88	0.03	**2.20**	**2.28**	0.91
Step-up[2]	**1.20**[2]	0.53	0.03	**1.25**[2]	0.66	0.03
Kneel	0.46	0.05	0.05	0.49	0.19	0.04
Lunge[3]	0.00	0.68	0.03	**2.56**[3]	**2.97**[3]	**2.38**[3]

[1]Medial lift-off and separation consistent with the force data in Table 6.1.
[2]Medial and lateral lift-off, disagreeing with separation and with the forces in Table 6.1.
[3]Lateral lift-off and separation, disagreeing with the force data in Table 6.1.

Table 6.2: Maximal lift-off, separation (Standard) and separation (Corrected) for the medial and lateral condyle for all five tasks: heelstrike, toe-off step-up, lunge and kneel. Note from the results of phantom experiment 1 that lift-off is not influenced by the error in the out-of-plane position. Values in bold are above the lift-off threshold of 1.0*mm.*

The mean, standard deviation and minimum force, medially and laterally are presented in Table 6.1. In addition the medial portion is presented as a percentage of the total force in the same table. Table 6.2 presents the lift-off and separations measured using pose estimation on the same in vivo datasets.

Heelstrike and Toe-off: The heelstrike and toe-off force data showed minimum medial contact forces of 39N and 11N respectively and minimum lateral forces of 87N and 76N. The medial values were below 50N, suggesting that there could have been a few instances with actual loss of contact.

Fluoroscopic heelstrike data shows maximal lift-off on the medial side of 5.11mm with corresponding separation of 2.24mm and 2.6mm for the *Standard* and *Corrected* methods. Laterally, similar lift-off and separation values were found, ranging from 4.04mm to 5.06mm for the three methods. The fluoroscopic toe-off data shows potential lift-off on the lateral side with 2.20mm lift-off and 2.28mm for the *Standard* method and 0.91mm for the *Corrected* method.

Thus, both force data and lift-off / separation data indicate that loss of contact is possible in both the heelstrike and toe-off data.

Step-up: The force data is distributed $54 - 46\%$ medial-lateral over the step-up, with a small standard deviation of 7%. Even at the minimum (236N or 206N) there is still sufficient contact force to rule out loss of contact. The fluoroscopic data for the step-up motion presents lift-off values larger than the threshold of 1.0mm (medial: 1.2mm, lateral 1.25mm), but separation stays well below 1.0mm for the *Standard* and *Corrected* methods. We attribute these separation values to measurement noise.

Therefore, the force data and the separation measurement show that no loss of contact has occurred during the step-up, whereas the lift-off data showed loss of contact.

Kneel: The kneeling force data demonstrates a consistent relative contact force on the medial side with a mean of 57%, with a standard deviation of 3%. The fluoroscopic data for the kneel task shows no hint of contact loss with lift-off and separations remaining below 0.5mm.

In the kneel data, both force data and lift-off / separation data indicate that no loss-of contact has occurred.

Lunge: The lunge task showed the most interesting discrepancy between force and fluoroscopic data (Figure 6.2). Forces on the medial side accounted for 58% of the total force with negligible standard deviation, indicating that no contact loss had occurred during the lunge motion. Fluoroscopic data showed large lateral lift-off values of 2.56mm, with similar separation values (*Standard*: 2.56mm, *Corrected*: 2.38mm).

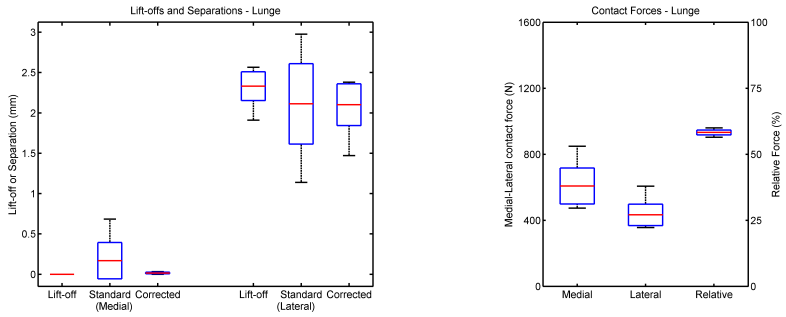


Figure 6.2: Boxplots for the fluoroscopic lift-off and separation (both with the Standard method and the Corrected method) on the left side and the contact force on the right side for the Lunge dataset.

The force data showed that no loss of contact had occurred during the lunge motion, while both the lift-off and the separation measures were well above the threshold for measuring contact loss, indicating lateral condylar contact loss.

Phantom Experiment 1

Figure 6.3 presents medial and lateral lift-off and separation. Clearly visible is the similarity in the shape of the profiles, but also the large amount of variation in some regions. Medially, the mean difference between lift-off and separation was 0.20mm ($p < 0.001$) with a standard deviation of 0.35mm and maximum value of 0.82mm. Laterally, the difference was -0.24mm ($p < 0.001$) with a standard deviation of 0.30mm and maximum of 1.04mm. With lift-off thresholds in the literature ranging from 0.51.0mm, the maximum difference of 1.04mm and 0.82mm indicates that there is a discrepancy between lift-off and separation.

Phantom Experiment 2

The two pose estimation methods showed the same differences between lift-off and stereo separation (Table 6.3): 0.22mm $\pm$ 0.36mm medially and -0.25mm $\pm$ 0.29mm laterally.

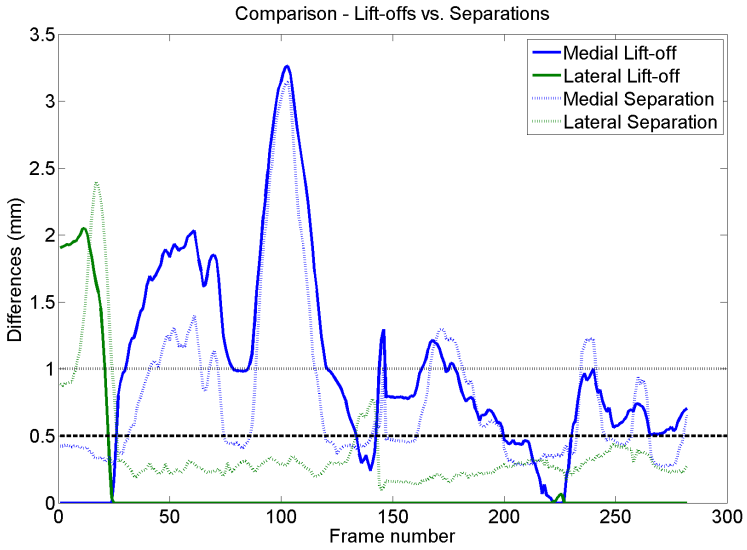


Figure 6.3: Lift-off and Separation assessed using the estimated prosthesis poses from a stereo phantom measurement. The solid lines indicate the lift-offs over all frames, whereas the dotted lines indicate the corresponding separation. Note how the medial lift-off shows a similar shape as the separation, but with large differences in many frames. The black dashed lines indicate the thresholds of 0.5 and 1.0mm often used when distinguishing lift-off from contact loss.

On the other hand, the separation measure demonstrated a little more variation in its differences with respect to the *Bi-plane separation reference* method: medially, the *Standard* method overestimated the amount of separation by 0.22mm ($p < 0.001$), while the *Corrected* method underestimated it by -0.20mm ($p < 0.001$). The standard deviations were relatively small for single-plane fluoroscopy: 0.35mm and 0.29mm for the *Standard* and *Corrected* methods, respectively.

Laterally, similar results were obtained: with the *Standard* method showing a difference in separation of 0.21mm with respect to the *Bi-plane separation reference* method with a standard deviation of 0.49mm. The *Corrected* method showed a difference of -0.31mm and a standard deviation of 0.29mm.

	Medial (mm)		Lateral (mm) (%)	
Lift-off	Mean	Std	Mean	Std
Standard method	0.22	0.36	−0.25	0.29
Corrected method	0.22	0.36	−0.25	0.29
Separation				
Standard method	0.22	0.35	0.21	0.49
Corrected method	−0.20	0.29	−0.31	0.29

Table 6.3: Differences with respect to standard separation from bi-plane data for lift-off and separation as measured on single-plane data.

6.6 Discussion

Condylar lift-off as measured with single-plane fluoroscopic analysis has been reported in the literature as a surrogate measure for contact loss [Stiehl et al., 1999, Dennis et al., 2001, Insall et al., 2002]. Separation was introduced later as a more accurate measure for actual contact loss [Kanekasu et al., 2004]. In this study, we compared lift-off and separation as measured from single plane fluoroscopy with a gold standard measurement using in-vivo data, as well as phantom data.

The in vivo fluoroscopic and force data present a mixed outcome. For kneeling and gait activities, the force data and lift-off or separation data are consistent. In kneeling, both the force data and the single-plane fluoroscopy suggest that no loss of contact occurred. In gait, the measured forces were below the 50N threshold, suggesting contact loss may have been possible. However, with 12N and 3N measured, it is still feasible that there was contact with low loads and our data do not allow us to discriminate this from loss of contact.

Lift-off during gait has been reported before in the literature [Stiehl et al., 1999, Dennis et al., 2003] with lift-off values of several millimeters, comparable to these results. However, our results demonstrate a possible mismatch of fluoroscopic lift-off values with actual force measurements. Actual loss of contact during gait has not been verified with measurements other than fluoroscopy, but it has been predicted in a single emg-driven model study [Kumar et al., 2013].

However, the stair and lunge data were inconsistent. The stair forces show no loss of contact, consistent with the separation measures, but lift-off was demonstrated. The forces measured during the lunge activity indicate no loss of contact, while the single-plane fluoroscopic measures indicate lateral condylar liftoff (2.5mm) and separation (2.5mm). There are two possible explanations for the discrepancies in the lunge data: First, single-plane fluoroscopy is not a sufficiently accurate basis for measuring condylar contact loss. Second, the lunge images show the posterior femoral cortex could be in contact with the tibial insert, creating a posterior impingement that transmits load while the lateral femoral condyle is not touching the tibial articular surface. The data available do not allow further discrimination of these two possibilities.

The results from phantom experiment 1 demonstrate that lift-off is not equivalent to separation, because with separation (as the closest distance between insert and condyle) the insert geometry is taken into account. With an accurate bi-plane measurement, separation can be considered an accurate measure of contact loss. With lift-off, the relation with the insert is not used at all and this can perhaps explain the measured values: the difference between lift-off and separation can be as high as 1mm (see the first 20 frames for the lateral side in Figure 6.3). Medially, there was a difference of approximately 0.5mm throughout large parts of the dataset.

The results from phantom experiment 2 demonstrate a possible effect of error in the out-of-plane position on separation, but not on lift-off. The definition of lift-off indicates that it is only influenced by the relative orientation of the femoral component and hence there is no effect of the out-of-plane position. Separation, however, requires an accurate femoral and insert position, demonstrated by the differences in separation between the two pose estimation methods: The *Standard* method can overestimate the amount of separation (by 0.22mm), while the *Corrected* method can underestimate it (by -0.20mm). These over- and underestimations show that the separation measure is not necessarily a good alternative for lift-off, when determined with single-plane fluoroscopy.

In our phantom experiment we had a bi-plane fluoroscopic setup, with digital

image detectors, yielding high quality images (high resolution, frame rate and contrast). This explains the relatively low standard deviations for lift-off and separation when compared to the in vivo data.

In the in vivo data, the image quality was considerably lower, resulting in a lower accuracy for pose estimation. We assume that this accounts for some of the discrepancies: lower image quality causes larger measurement errors in the femoral and tibial orientations, in turn yielding higher lift-off and separation values. Nonetheless, lower image quality is generally expected in clinical data and we are certain that similar discrepancies between in vivo force and lift-off or separation are possible in other clinical data. Especially when measuring lift-off occurrences using a threshold, it is likely that due to larger measurement errors some frames will demonstrate lift-off above the threshold. Another limitation was that good frame-by-frame synchronization was not available in the in vivo data. With synchronization, it would have been possible to investigate the exact relation between implant pose and measurement error in lift-off. Instead, we compared overall minimum and mean in vivo force to overall minimum and mean lift-off.

At a minimum, the discrepancy between high condylar forces and observed lift-off in the step activity suggest liftoff should not be used as an indication of condylar loss of contact. The lunge data present a similar discrepancy suggesting neither lift-off nor separation from single-plane fluoroscopy are reliable, but the possibility of posterior impingement weakens that conclusion. Another limitation of the study was that there was no quantitative measure, which could have convincingly said that there was loss of contact. Consequently a threshold of 50N was chosen to detect contact safely.

Furthermore, only a single prosthesis type was used which makes it difficult to extrapolate the results to other prosthesis types. Unfortunately, only a few instrumented knee designs exist, making it difficult to retrieve contact force data for other implants.

We conclude that lift-off and separation as measured with single-plane fluoroscopic analysis can lead to an overestimation of the magnitude and incidence of actual contact loss between the femoral component and tibial insert. If used, the separation measurement is shown to be a better indi-

cator of contact loss, but should be reported with appropriate statistical confidence levels corresponding to the imaging and activity details of each study.

We do not dispute the possibility of contact loss after TKA and its possible effect on the wear of the polyethylene insert. We do, however, recommend taking great care when drawing conclusions on contact loss based on single plane fluoroscopic analysis and a lift-off threshold of 1.0mm. Higher lift-off thresholds may be more reliable, with the risk of not detecting condylar lift-off while it actually occurs. To further explore this, an in vivo bi-plane fluoroscopic experiment is needed, with synchronized measurements of internal contact forces from instrumented knee prostheses.

6.6.1 Acknowledgement

We sincerely thank Darryl D'Lima, Cliff Colwell and BJ Fregly for sharing their eKnee data with us.

CHAPTER 7

DISCUSSION AND RECOMMENDATIONS

In the introduction of this thesis, several limitations were identified of model-based shape matching methods applied to both RSA and fluoroscopy. Both applications require an experienced user and the necessary manual interaction during the analysis makes the results operator-dependent and time-consuming. The aim of this research was to improve the reliability and usability of model-based shape matching for RSA and fluoroscopic analysis. In this chapter these improvements will be presented and discussed and recommendations for future research directions will be given.

The reliability, in terms of accuracy and robustness, of model-based shape matching is a recurring subject in all chapters in this thesis. Errors in model-based shape matching for RSA (Chapter 3) are lower than 0.2 mm and 1°, which are in the same range as the accuracy commonly reported in the literature [Kaptein et al., 2006, Mjöberg et al., 1986, Kärrholm, 1989, Kärrholm et al., 1994, Börlin et al., 2002]. For fluoroscopy the errors are below 0.5 mm for the in-plane positions and below 0.5° for rotations (Chapters 2 and 4). This is comparable to results in the literature, where the reported errors range from 0.09 mm to 0.40 mm for in-plane positions and from 0.35° to 1.3° for orientations [Banks and Hodge, 1996, Hoff et al., 1998, Mahfouz et al., 2003, Komistek et al., 2003, Kanisawa et al., 2003, Garling et al., 2005].

The accuracy of fluoroscopy is considered sufficient for most joint kinematics studies [Stiehl et al., 1995, Banks and Hodge, 1996, Zuffi et al., 1999, Mahfouz et al., 2003, Komistek et al., 2003]. For example, fluoroscopy is used to measure condylar lift-off in total knee arthroplasty [Stiehl et al., 1999, Dennis et al., 2001, Insall et al., 2002]. A difference of 0.5–1.0 mm or larger between the medial and lateral minimum joint space width is considered to indicate loss of contact [Stiehl et al., 1999, Dennis et al., 2001]. However, in Chapter 6, a difference of 5.1 mm was found with single-plane fluoroscopy, although the in vivo contact force from an instrumented knee implant still showed contact. This indicates that the commonly reported accuracies for single-plane fluoroscopy are still not sufficient to measure condylar loss of contact.

The robustness of automated pose estimation methods for model-based shape matching was assessed in Chapter 5. The automated pose estima-

tion presented erroneous pose estimates in 15% of the frames in clinical fluoroscopy data of the ROCC total knee implant. For example, on a hypothetical small dataset of twenty frames, three frames would require manual corrections to the pose and a restart of the optimization. This makes the method cumbersome to use and it indicates that supervision and manual intervention remain important to get accurate measurements from fluoroscopic analysis.

Similar failure-rates were presented by Mahfouz et al. [2003] ranging from 5% to 50% depending on the initial starting pose provided by the user before automated pose estimation. It was discussed that this high failure-rate was due to the symmetrical shape of the implant. A more recent paper presented a shape matching method, which simultaneously estimates the pose for two symmetrical solutions, but it did not report quantitative results on the impact of the symmetry on the accuracy or the likelihood of success [Hermans et al., 2007].

The specific shape of an implant will also have an effect on the accuracy of model-based shape matching. The virtual projection of an implant has to be sufficiently unique and distinctive for each pose of the implant, such that it can be matched in similarity with the detected implant contour. Rotating a hip stem about its longitudinal axis results in no or minor changes to its silhouette due to its near cylindrical shape. This makes the measurement of longitudinal rotation inaccurate (Chapter 3). The addition of its spherical head in the shape matching resulted in improved accuracy and precision, because the head has an offset with respect to the longitudinal stem and introduces additional variation to the silhouette.

Besides the shape of the implant, the effect of the accuracy of the three-dimensional implant model on the shape matching has been reported earlier by our group: the use of reverse engineered (RE) models increases the accuracy of the pose measurements when compared to computer aided design (CAD) models [Kaptein et al., 2003].

The detection of the outer contour of the implant silhouette is often a manual or semi-automatic procedure, utilizing basic image processing techniques [Banks and Hodge, 1996, Zuffi et al., 1999, Kaptein et al., 2004]. In older

fluoroscopy systems the quality of the images can be poor: low resolution, low contrast and severe image deformation. High frame rates are necessary for dynamic analysis or analysis of high-speed motion. However, to limit X-ray dose for the patient the dose per frame is reduced leading to low contrast images. In addition, implant motion introduces motion blur, reducing the accuracy of the detected implant silhouette.

Fregly et al. [2005] have presented the effects of X-ray attenuation on the accuracy of the silhouette in the image and how a biased edge detection procedure could introduce errors up to 2 mm and 2° in the pose measurements. Mahfouz et al. [2005] reported similar errors and even recommended a pose estimation process without "manual, a priori segmentation".

In Chapter 2, a shape matching approach is presented, which eliminates such an "a priori segmentation" by integrating the contour detection and pose estimation. On good quality images, this method allows for a fully automated fluoroscopic analysis. On lower quality image data it was demonstrated that such an approach eliminates large portions of the cumbersome manual contour detection. This resulted in a threefold increase in efficiency while keeping sufficient precision for clinical research: within 1 mm and 1°.

This research has focused on improving the accuracy of single plane fluoroscopy, as this is the common fluoroscopic setup used in kinematics studies [Banks and Hodge, 1996, Hoff et al., 1998, Zuffi et al., 1999, Komistek et al., 2003, Kanisawa et al., 2003, Garling et al., 2005, Mahfouz et al., 2003, Li et al., 2008]. In single plane analysis, the accuracy in the out-of-plane direction is relatively low or sometimes not even reported [Banks and Hodge, 1996, Hoff et al., 1998, Mahfouz et al., 2003, Komistek et al., 2003, Kanisawa et al., 2003]. The entire knee implant was considered as a single model with twelve degrees of freedom instead of separate models for the femur and tibia implant components with six degrees of freedom each (Chapter 4). By adding a collision constraint, physically impossible intersections between the femoral component and the polyethylene insert were prevented. This greatly improved the accuracy of single-plane analysis from 2.0 ± 0.7 mm to 0.1 ± 0.1 mm in the out-of-plane position.

Although bi-plane fluoroscopy setups eliminate the large out-of-plane error and increases the overall accuracy, they remains impractical to serve as an alternative to single-plane fluoroscopy. They are not widely available and the space for the movement task is limited. As a result, these systems are only used in some highly-specialized motion laboratories [Tashman and Anderst, 2003, You et al., 2001, Li et al., 2008].

In conclusion, the reliability of model-based shape matching has been improved for both the analysis of hip stems in RSA and the analysis of total knee implants in fluoroscopy. A study was presented on the success-rate of automated pose estimation methods and improvements to the contour detection and optimization methods were presented, which greatly increased their usability.

7.1 Recommendations

Several improvements to the usability and reliability were shown in Chapters 2 – 6, but other interesting areas of research were also identified as future work together with the following recommendations.

The validation of single-plane fluoroscopic analysis is problematic on actual patient data, because it is impossible to setup a non-invasive reference measurement with patients. Instead most studies have performed phantom studies. In addition it is difficult to extrapolate the results from phantom data to actual clinical practice. Unfortunately, a fair comparison with other validation studies is difficult due to differences in experimental setup and testing methodologies. Fluoroscopic setups range from older single-plane setups with low contrast and severe image deformation to bi-plane modern flat-panel systems, which can reach higher frame rates.

Therefore, it seems prudent to standardize the validation of new methods for fluoroscopic analysis and shape matching, similar to the standardization of RSA methods [Valstar et al., 2005, ISO, 2013]. In this thesis for example a bi-plane fluoroscopic setup was used as reference standard for validation of model-based fluoroscopy (Chapters 2 and 4). This bi-plane setup yielded

high quality images and shape matching on both bi-plane data yields highly accurate reference pose measurements of the implant. The performance of single-plane fluoroscopic analysis was investigated by discarding the images with the anterior-posterior viewpoint, effectively simulating a single-plane setup.

Instead of standardizing fluoroscopic results, it could be useful if a public reference dataset would be available of (phantom) fluoroscopic data with known, calibrated focus positions, implant model and known reference poses. New methods for model-based fluoroscopic analysis should be validated on this standard reference dataset, so that a fair comparison can be made between the results of various methods on the same data.

A model-based shape matching method was presented that eliminates the manual contour detection (Chapter 2). Unfortunately, the method still requires at least ten seconds on a single fluoroscopic frame due to the optimization. A failure rate of 15% in clinical data was presented in Chapter 5, which indicates that unsupervised automatic analysis is infeasible with the currently applied methods.

In Chapter 2 it was discussed how fully automatic analysis may be feasible by propagating the results from one frame to the next. As future work, this should be extended to a multi-frame analysis measuring the implant motion over a series of frames instead of measuring pose in each frame separately. Improvements to the optimization were discussed in Chapter 5: combining the results of several optimization methods or detecting unsuccessful optimizations and automatically restarting, in order to improve the success-rate. With these improvements (Chapter 5 and Chapter 2), it may be feasible to develop a shape matching method, which can estimate the implant pose in less than a second with fast and robust optimization and with automatic contour detection, which will make the method more intuitive and interactive to use.

In this thesis various factors were explored which influence the reliability of model-based shape matching for RSA and fluoroscopy and improvements to the usability were demonstrated by reducing the amount of manual work. In order to get robust, automated analysis, the quality and resolution of

fluoroscopic frames needs to improve. Fortunately, as the technology of fluoroscopy advances, high quality data will be more easily available. But with high-quality data, the amount of data produced with fluoroscopic analysis is also getting larger, and may soon grow beyond the capabilities of manual, supervised pose estimation methods. It is therefore imperative that fully automatic unsupervised methods are developed.

CHAPTER 8

SUMMARY

Model-based shape matching is a tool commonly used for clinical research, for example to measure the migration of an implant with respect to the bone over several years with Roentgen stereophotogrammetric analysis (RSA) or to validate the design and functioning of knee prostheses by measuring implant kinematics with fluoroscopy. The aim of this thesis was to investigate practical issues limiting the general usability of shape matching and to develop solutions, which improve the reliability of shape matching for RSA and fluoroscopy.

A significant part of the workflow of model-based shape matching is the manual or semi-automatic segmentation of the implant contour from its silhouette in a radiograph. In Chapter 2 an automatic contour detection approach was proposed, which integrates contour detection directly into the pose estimation. This has greatly reduced the amount and complexity of the work required on low quality fluoroscopic data.

The methods presented in the next two Chapters (3 and 4) improved the accuracy and precision of shape matching by combining multiple implant parts into a single model. The pose estimation was performed on the whole model, correctly taking the relation between parts into account. Both approaches reduced the amount of work otherwise needed when the shapes of the components are matched separately. In Chapter 3 the measurement of the longitudinal rotation was improved from 0.96° to 0.69° for a hip stem. In Chapter 4 the out-of-plane error of single-plane fluoroscopic analysis was reduced from several millimeters to submillimeter accuracy, sufficiently accurate for most clinical research purposes.

The robustness of model-based shape matching was investigated in Chapter 5 where it was applied fully automatically on fluoroscopic data. On clinical data, only 15% of the frames required manual supervision. The optimal success-rate on phantom data (97%) indicated that even less supervision may become feasible, but that some manual corrections remain necessary to get accurate results from fluoroscopic analysis.

In chapter 6, the potential consequences were investigated of interpreting data from fluoroscopic analyses when the method has only a limited accuracy. A large discrepancy was demonstrated for both lift-off and separation,

as measured with single-plane fluoroscopy, compared to the actual loss of contact as measured with in instrumented knee implant with force sensors.

In conclusion, several improvements were made to the usability and reliability of model-based shape matching for RSA and fluoroscopy. There is still room for more improvement in the areas of usability and automated analysis. The standardization of validating new fluoroscopic methods could be a welcome addition to the biomechanical field, making it a lot easier to compare results between methods and experiments. As the technology of fluoroscopy advances, the amount of high quality data is getting larger, and may soon grow beyond the capabilities of manual, supervised pose estimation methods. It is therefore imperative that fully automatic unsupervised methods are developed.

CHAPTER 9

SAMENVATTING

Model-based shape matching wordt vaak toegepast binnen klinisch onderzoek, bijvoorbeeld als tool om de migratie van een prothese ten opzichte van het bot over meerdere jaren te meten met Röntgen stereofotogrammetrische analyse (RSA), of bijvoorbeeld om het ontwerp en de functie van een knie prothese te valideren door de prothesekinematica te meten met fluoroscopie. Het doel van dit proefschrift was om de praktische problemen te onderzoeken die de algemene bruikbaarheid van shape matching beperken en om nieuwe oplossingen aan te dragen waarmee de betrouwbaarheid van shape matching voor RSA en fluoroscopie verbeterd kon worden.

Een groot deel van het werk met model-based shape matching bestaat uit het handmatig of semiautomatisch segmenteren van de buitenste contour van het silhouette van de prothese in een Röntgenbeeld. In Hoofdstuk 2 werd een automatische aanpak geïntroduceerd waar de contourdetectie is geïntegreerd met het bepalen van de positie en oriëntatie. Hierdoor is de hoeveelheid handwerk en de complexiteit in het analyseren van lage kwaliteit Röntgenbeelden duidelijk minder geworden.

De methodes in de volgende twee Hoofdstukken (3 en 4) hebben de nauwkeurigheid en precisie verbeterd van shape matching door protheseonderdelen te combineren in één model. Het bepalen van de positie en oriëntatie werd gedaan met het gehele model, waarbij de relatie tussen de onderdelen werd meegenomen. Beide aanpakken vereisten minder handwerk ten opzichte van een aanpak waar shape matching voor elk onderdeel afzonderlijk wordt gedaan. In Hoofdstuk 3 werd de meetfout verlaagd van 0.96° naar 0.69° van de rotatie rond de lengte as van een heupsteel. In Hoofdstuk 4 werd de meetfout verlaagd van meerdere millimeters tot submillimeter nauwkeurigheid voor de positie uit het beeldvlak met monofluoroscopie, waarmee de methode voldoende nauwkeurig is geworden voor het meeste klinische onderzoek.

De robuustheid van model-based shape matching werd onderzocht in Hoofdstuk 5 waarbij de methode volledig automatisch werd toegepast op fluoroscopiedata. Slechts 15% van de Röntgenbeelden had verdere supervisie nodig in de klinische data. De optimale kans van slagen op de fantoom data (97%) gaf aan dat verdere verbeteringen haalbaar zijn, maar tegelijk-

ertijd dat enkele handmatige correcties noodzakelijk zullen blijven voor een nauwkeurige fluoroscopieanalyse.

In hoofdstuk 6 zijn de mogelijke consequenties onderzocht wanneer fluoroscopiedata wordt geïnterpreteerd terwijl de methode slechts een beperkte nauwkeurigheid heeft. Een groot verschil werd gedemonstreerd voor zowel lift-off als separation, gemeten met monofluoroscopie, vergeleken met het werkelijke contact, gemeten met een knieprothese geïnstrumenteerd met ingebouwde krachtsensoren.

In dit proefschrift zijn verschillende verbeteringen getoond in de bruikbaarheid en betrouwbaarheid van model-based shape matching voor RSA en fluoroscopie. Er is echter nog meer ruimte voor verbetering en het automatiseren van de analyse. Een standaardisatie voor het valideren van nieuwe methodes voor fluoroscopie zou een goede toevoeging zijn aan het vakgebied, waarmee het een stuk eenvoudiger moet worden om resultaten te vergelijken tussen shape matching methodes en experimenten. Omdat de fluoroscopietechniek zich verder ontwikkeld, zal de hoeveelheid hoge kwaliteit data steeds groter worden. Hierdoor zal het steeds moeilijker worden om deze data met de hand te verwerken. Het is dan ook essentieel dat volledig automatische methodes worden ontwikkeld.

REFERENCES

Implants for surgery – roentgen stereophotogrammetric analysis for the assessment of migration of orthopaedic implants. Standard, International Organization for Standardization, Geneva, CH, 2013.

Thomas P Andriacchi. Dynamics of knee malalignment. *The Orthopedic clinics of North America*, 25(3):395–403, 1994.

H. Baldursson, N. Egund, L.I. Hansson, and G. Selvik. Instability and wear of total hip prostheses determined with roentgen stereophotogrammetry. *Archives of Orthopaedic and Trauma Surgery*, 95(4):257–263, 1979.

S.A. Banks and W.A. Hodge. Accurate measurement of three-dimensional knee replacement kinematics using single-plane fluoroscopy. *Biomedical Engineering, IEEE Transactions on*, 43(6):638–649, 1996.

Scott A Banks, MK Harman, Johan Bellemans, and WA Hodge. Making sense of knee arthroplasty kinematics: news you can use. *The Journal of Bone & Joint Surgery*, 85(suppl_4):64–72, 2003.

Arnaud Barre, J Thiran, B Jolles, Nicolas Theumann, and Kamiar Aminian. Soft tissue artifact assessment during treadmill walking in subjects with total knee arthroplasty. 2013.

R.E. Bellman and S.E. Dreyfus. *Applied dynamic programming*, volume 7962. Princeton University Press, 1966.

N Börlin, T Thien, and J Kärrholm. The precision of radiostereometric measurements. manual vs, digital measurements. *Journal of Biomechanics*, 35(1):69–79, 2002.

J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

F Catani, S Fantozzi, A Ensini, A Leardini, D Moschella, and S Giannini. Influence of tibial component posterior slope on in vivo knee kinematics in fixed-bearing total knee arthroplasty. *Journal of orthopaedic research*, 24(4):581–587, 2006.

A Andrew R Conn, Katya Scheinberg, and Luis N Vicente. *Introduction to derivative-free optimization*, volume 8. Siam, 2009.

Patrick A Costigan, Kevin J Deluzio, and Urs P Wyss. Knee and hip kinetics during normal stair climbing. *Gait & posture*, 16(1):31–37, 2002.

D.A. Dennis, R.D. Komistek, W.A. Hoff, and S.M. Gabriel. In vivo knee kinematics derived using an inverse perspective technique. *Clinical orthopaedics and related research*, 331:107–117, 1996.

DA Dennis, RD Komistek, SA Walker, EJ Cheal, and JB Stiehl. Femoral condylar lift-off in vivo in total knee arthroplasty. *Journal of Bone & Joint Surgery, British Volume*, 83(1):33–39, 2001.

Douglas A Dennis, Richard D Komistek, and Mohamed R Mahfouz. In vivo fluoroscopic analysis of fixed-bearing total knee replacements. *Clinical orthopaedics and related research*, 410:114–130, 2003.

Darryl D DLima, Christopher P Townsend, Steven W Arms, Beverly A Morris, and Clifford W Colwell Jr. An implantable telemetry device to measure intra-articular tibial forces. *Journal of biomechanics*, 38(2):299–304, 2005.

Brian P Flannery, William H Press, Saul A Teukolsky, and William Vetterling. Numerical recipes in c. *Press Syndicate of the University of Cambridge, New York*, 1992.

B.J. Fregly, H.A. Rahman, and S.A. Banks. Theoretical accuracy of model-based shape matching for measuring natural knee kinematics with single-plane fluoroscopy. *Journal of biomechanical engineering*, 127(4):692, 2005.

EH Garling, BL Kaptein, B Mertens, W Barendregt, HEJ Veeger, RGHH Nelissen, and ER Valstar. Soft-tissue artefact assessment during step-up using fluoroscopy and skin-mounted markers. *Journal of Biomechanics*, 40:S18–S24, 2007.

E.H.E.H. Garling, B.L.B.L. Kaptein, K. Geleijns, R.R.G.H.H. Nelissen, and E.R.E.R. Valstar. Marker configuration model-based roentgen fluoroscopic analysis. *Journal of biomechanics*, 38(4):893–901, 2005.

Gurdev S Gill, Atul B Joshi, and David M Mills. Total condylar knee arthroplasty: 16-to 21-year results. *Clinical orthopaedics and related research*, 367:210–215, 1999.

HS Gill, J Alfaro-Adrián, C Alfaro-Adrián, P McLardy-Smith, and DW Murray. The effect of anteversion on femoral component stability assessed by radiostereometric analysis. *The Journal of Arthroplasty*, 17 (8):997–1005, 2002.

G.R. Hanson, J.F. Suggs, A.A. Freiberg, S. Durbhakula, and G. Li. Investigation of in vivo 6dof total knee arthoplasty kinematics using a dual orthogonal fluoroscopic system. *Journal of orthopaedic research*, 24(5): 974–981, 2006.

Leif I Havelin, Anne M Fenstad, Roger Salomonsson, Frank Mehnert, Ove Furnes, Søren Overgaard, Alma B Pedersen, Peter Herberts, Johan Kärrholm, and Göran Garellick. The nordic arthroplasty register association: a unique collaboration between 3 national hip arthroplasty registries with 280,201 thrs. *Acta orthopaedica*, 80(4):393–401, 2009.

B. Heinlein, F. Graichen, A. Bender, A. Rohlmann, and G. Bergmann. Design, calibration and pre-clinical testing of an instrumented tibial tray. *Journal of biomechanics*, 40:S4–S10, 2007.

Jeroen Hermans, Johan Bellemans, Dirk Vandermeulen, and Paul Suetens. A statistical approach to determine symmetrical solutions for the registration of 3d knee implant models to sagittal fluoroscopy images. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8. IEEE, 2007.

Jeroen Hermans, Johan Bellemans, Frederik Maes, Dirk Vandermeulen, and Paul Suetens. A statistical framework for the registration of 3d knee implant components to single-plane x-ray images. In *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on*, pages 1–8. IEEE, 2008.

S. Hirokawa, M. Abrar Hossain, Y. Kihara, and S. Ariyoshi. A 3d kinematic estimation of knee prosthesis using x-ray projection images: clinical assessment of the improved algorithm for fluoroscopy images. *Medical and Biological Engineering and Computing*, 46(12):1253–1262, 2008.

W.A. Hoff, R.D. Komistek, D.A. Dennis, S. Walker, E. Northcut, and K. Spargo. Pose estimation of artificial knee implants in fluoroscopy images using a template matching technique. In *Applications of Computer Vision, 1996. WACV'96., Proceedings 3rd IEEE Workshop on*, pages 181–186. IEEE, 1996.

W.A. Hoff, R.D. Komistek, D.A. Dennis, S.M. Gabriel, S.A. Walker, et al. Three-dimensional determination of femoral-tibial contact positions under in vivo conditions using fluoroscopy. *Clinical Biomechanics*, 13(7):455–472, 1998.

J.N. Insall, G.R. Scuderi, R.D. Komistek, K. Math, D.A. Dennis, and D.T. Anderson. Correlation between condylar lift-off and femoral component alignment. *Clinical orthopaedics and related research*, 403:143, 2002.

K. Kanekasu, S.A. Banks, S. Honjo, O. Nakata, and H. Kato. Fluoroscopic analysis of knee arthroplasty kinematics during deep flexion kneeling. *The Journal of arthroplasty*, 19(8):998–1003, 2004.

I. Kanisawa, A.Z. Banks, S.A. Banks, H. Moriya, and A. Tsuchiya. Weight-bearing knee kinematics in subjects with two types of anterior cruciate ligament reconstructions. *Knee Surgery, Sports Traumatology, Arthroscopy*, 11(1):16–22, 2003.

Bart L Kaptein, Kevin B Shelburne, Michael R Torry, and J Erik Giphart. A comparison of calibration methods for stereo fluoroscopic imaging systems. *Journal of biomechanics*, 44(13):2511–2515, 2011.

BL Kaptein, ER Valstar, BC Stoel, PM Rozing, and JHC Reiber. A new model-based rsa method validated using cad models and models from reversed engineering. *Journal of biomechanics*, 36(6):873–882, 2003.

BL Kaptein, ER Valstar, BC Stoel, PM Rozing, and JHC Reiber. Evaluation of three pose estimation algorithms for model-based roentgen stereophotogrammetric analysis. *Proceedings of the Institution of Mechanical Engineers*, 218(4):231–238, 2004.

BL Kaptein, Valstar ER, CW Spoor, BC Stoel, and Rozing PM. Model-based rsa of a femoral hip stem using surface and geometric shape models. *Clinical Orthopaedics and Related Research*, 448:92–97, 2006.

J Kärrholm. Roentgen stereophotogrammetry. review of orthopedic applications. *Acta orthopaedica Scandinavica*, 60(4):491–503, 1989.

J Kärrholm, B Borssén, G Löwenhielm, and F Snorrason. Does early micromotion of femoral stem prostheses matter? 4-7-year stereoradiographic follow-up of 84 cemented prostheses. *Journal of Bone & Joint Surgery, British Volume*, 76(6):912–7, 1994.

J Kärrholm, P Herberts, P Hultmark, H Malchau, B Nivbrant, and J Thanner. Radiostereometry of hip prostheses: review of methodology and clinical results. *Clinical Orthopaedics and Related Research*, pages 94–110, 1997.

Kenton R Kaufman, Nebojsa Kovacevic, Steven E Irby, and Clifford W Colwell. Instrumented implant for measuring tibiofemoral forces. *Journal of biomechanics*, 29(5):667–671, 1996.

Young-Hoo Kim, Hee-Kyun Kook, and Jun-Shik Kim. Comparison of fixed-bearing and mobile-bearing total knee arthroplasties. *Clinical orthopaedics and related research*, 392:101–115, 2001.

S. Kirkpatrick. Optimization by simulated annealing: Quantitative studies. *Journal of statistical physics*, 34(5):975–986, 1984.

R.D. Komistek, D.A. Dennis, and M. Mahfouz. In vivo fluoroscopic analysis of the normal human knee. *Clinical orthopaedics and related research*, 410: 69–81, 2003.

O.H.J. Koning, B.L. Kaptein, E.H. Garling, J.W. Hinnen, J.F. Hamming, E.R. Valstar, and JH van Bockel. Assessment of three-dimensional stent-graft dynamics by using fluoroscopic roentgenographic stereophotogrammetric analysis. *Journal of Vascular Surgery*, 46(4):773–779, 2007.

D Kumar, KT Manal, and KS Rudolph. Knee joint loading during gait in healthy controls and individuals with knee osteoarthritis. *Osteoarthritis and Cartilage*, 21(2):298–305, 2013.

S Lavallée and R Szeliski. Recovering the position and orientation of free-form objects from image contours using 3d distance maps. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(4):378–390, 1995.

Alberto Leardini, Lorenzo Chiari, Ugo Della Croce, Aurelio Cappozzo, et al. Human movement analysis using stereophotogrammetry. part 3. soft tissue artifact assessment and compensation. *Gait & posture*, 21(2):212, 2005.

Kenneth Levenberg. A method for the solution of certain problems in least squares. *Quarterly of applied mathematics*, 2:164–168, 1944.

G. Li, T.H. Wuerz, L.E. DeFrate, et al. Feasibility of using orthogonal fluoroscopic images to measure in vivo joint kinematics. *Journal of biomechanical engineering*, 126(2):314, 2004.

G. Li, S. Van de Velde, and J. Bingham. Validation of a non-invasive fluoroscopic imaging technique for the measurement of dynamic knee joint motion. *Journal of biomechanics*, 41(7):1616–1622, 2008.

M.I.A. Lourakis. levmar: Levenberg-marquardt nonlinear least squares algorithms in C/C++. [web page] `http://www.ics.forth.gr/~lourakis/levmar/`, Jul. 2004. [Accessed on 31 Jan. 2005.].

M.R. Mahfouz, W.A. Hoff, R.D. Komistek, and D.A. Dennis. A robust method for registration of three-dimensional knee implant models to two-dimensional fluoroscopy images. *Medical Imaging, IEEE Transactions on*, 22(12):1561–1574, 2003.

M.R. Mahfouz, W.A. Hoff, R.D. Komistek, D.A. Dennis, et al. Effect of segmentation errors on 3d-to-2d registration of implant models in x-ray images. *Journal of biomechanics*, 38(2):229–240, 2005.

Henrik Malchau, Johan Kärrholm, Yu Xing Wang, and Peter Herberts. Accuracy of migration analysis in hip arthroplasty digitized and conventional radiography, compared to radiostereometry in 51 patients. *Acta Orthopaedica*, 66(5):418–424, 1995.

Henrik Malchau, Peter Herberts, Thomas Eisler, Göran Garellick, and Peter Söderman. The swedish total hip replacement register. *The Journal of Bone & Joint Surgery*, 84(suppl_2):S2–S20, 2002.

Primoz Markelj, D Tomaževič, Bostjan Likar, and F Pernuš. A review of 3d/2d registration methods for image-guided interventions. *Medical image analysis*, 16(3):642–661, 2012.

Donald W Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial & Applied Mathematics*, 11(2):431–441, 1963.

J. Martin Bland and D.G. Altman. Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet*, 327 (8476):307–310, 1986.

B Mjöberg, G Selvik, LI Hansson, R Rosenqvist, and R Onnerfalt. Mechanical loosening of total hip prostheses. a radiographic and roentgen stereophotogrammetric study. *Journal of Bone & Joint Surgery, British Volume*, 68(5):770–774, 1986.

J.A. Nelder and R. Mead. A simplex method for function minimization. *The computer journal*, 7(4):308–313, 1965.

RGHH Nelissen. *Fundamental aspects of the clinical evaluation of total joint prostheses. The total condylar knee prosthesis.* PhD thesis, Leiden University, The Netherlands, 1995.

Rob GHH Nelissen, Edward R Valstar, and Piet M Rozing. The effect of hydroxyapatite on the micromotion of total knee prostheses. a prospective, randomized, double-blind study*. *The Journal of Bone & Joint Surgery*, 80(11):1665–72, 1998.

Rob GHH Nelissen, Edward R Valstar, Ruud G Pöll, Eric H Garling, and Ronald Brand. Factors associated with excessive migration in bone im-

paction hip revision surgery: a radiostereometric analysis study. *The Journal of arthroplasty*, 17(7):826–833, 2002.

Rob GHH Nelissen, Eric H Garling, and Edward R Valstar. Influence of cement viscosity and cement mantle thickness on migration of the exeter total hip prosthesis. *The Journal of arthroplasty*, 20(4):521–528, 2005.

Marc J Nieuwenhuijse, Edward R Valstar, Bart L Kaptein, and Rob GHH Nelissen. Good diagnostic performance of early migration as a predictor of late aseptic loosening of acetabular cupsresults from ten years of follow-up with roentgen stereophotogrammetric analysis (rsa). *The Journal of Bone & Joint Surgery*, 94(10):874–880, 2012.

Kjell G Nilsson and Johan Kärrholm. Increased varus-valgus tilting of screw-fixated knee prostheses: stereoradiographic study of uncemented versus cemented tibial components. *The Journal of arthroplasty*, 8(5):529–540, 1993.

L Nistor, D Blaha, U Kjellstöm, and G Selvik. In vivo measurements of relative motion between an uncemented femoral total hip component and the femur by roentgen stereophotogrammetric analysis. *Clinical Orthopaedics and Related Research*, 269:221–227, 1991.

I Önsten, K Åkesson, J Besjakov, and KJ Obrant. Migration of the Charnley stem in rheumatoid arthritis and osteoarthritis. A roentgen stereophotogrammetric study. *Journal of Bone & Joint Surgery, British Volume*, 77 (1):18–22, 1995.

K. Perlin. Noise hardware. *Real-Time Shading SIGGRAPH Course Notes*, 2001.

Bart G Pijls, Edward R Valstar, Klaas-Auke Nouta, Josepha WM Plevier, Marta Fiocco, Saskia Middeldorp, and Rob GHH Nelissen. Early migration of tibial components is associated with late revision: a systematic review and meta-analysis of 21,000 knee arthroplasties. *Acta orthopaedica*, 83(6):614–624, 2012.

AH Prins, BL Kaptein, BC Stoel, R. Nelissen, JHC Reiber, and ER Valstar. Handling modular hip implants in model-based rsa: Combined stem–head models. *Journal of biomechanics*, 41(14):2912–2917, 2008.

AH Prins, BL Kaptein, BC Stoel, JHC Reiber, and ER Valstar. Detecting femur–insert collisions to improve precision of fluoroscopic knee arthroplasty analysis. *Journal of biomechanics*, 43(4):694–700, 2010.

AD Reading, AW McCaskie, and PJ Gregg. The inadequacy of standard radiographs in detecting flaws in the cement mantle. *Journal of Bone & Joint Surgery, British Volume*, 81(1):167–170, 1999.

L Ryd, BE Albrektsson, L Carlsson, F Dansgard, P Herberts, A Lindstrand, L Regner, and S Toksvig-Larsen. Roentgen stereophotogrammetric analysis as a predictor of mechanical loosening of knee prostheses. *Journal of Bone & Joint Surgery, British Volume*, 77(3):377–383, 1995.

M Sati, JA De Guise, S Larouche, and G Drouin. Quantitative assessment of skin-bone movement at the knee. *The Knee*, 3(3):121–138, 1996.

PH Schönemann. A general solution to the orthogonal procrustes problem. *Psychometrika*, 31(1):1–10, 1966.

G Selvik. Roentgen stereophotogrammetry. a method for the study of the kinematics of the skeletal system. *Acta orthopaedica Scandinavica. Supplementum*, 232:1–51, 1989.

A. Sharma, R.D. Komistek, C.S. Ranawat, D.A. Dennis, and M.R. Mahfouz. In vivo contact pressures in total knee arthroplasty. *The Journal of arthroplasty*, 22(3):404–416, 2007.

I Söderkvist and PA Wedin. Determining the movements of the skeleton using well-configured markers. *Journal of Biomechanics*, 26(12):1473–1477, 1993.

Peter Spellucci. An sqp method for general nonlinear programs using only equality constrained subproblems. *Mathematical programming*, 82(3):413–448, 1998.

J.B. Stiehl, R.D. Komistek, D.A. Dennis, R.D. Paxson, and W.A. Hoff. Fluoroscopic analysis of kinematics after posterior-cruciate-retaining knee arthroplasty. *Journal of Bone & Joint Surgery, British Volume*, 77(6): 884–889, 1995.

J.B. Stiehl, D.A. Dennis, R.D. Komistek, and H.S. Crane. In vivo determination of condylar lift-off and screw-home in a mobile-bearing total knee arthroplasty. *The Journal of arthroplasty*, 14(3):293–299, 1999.

S. Tashman and W. Anderst. In-vivo measurement of dynamic joint motion using high speed biplane radiography and ct: application to canine acl deficiency. *Transactions-American Society of Mechanical Engineers, Journal of Biomechanical Engineering*, 125(2):238–245, 2003.

Pierre Terdiman. Opcode — optimized collision detection, the collision detection library from the open dynamics engine, 2001. URL `http://www.ode.org`.

Edward R Valstar, Richie Gill, Leif Ryd, Gunnar Flivik, Niclas Börlin, and Johan Kärrholm. Guidelines for standardization of radiostereometry (rsa) of implants. *Acta orthopaedica*, 76(4):563–572, 2005.

ER Valstar. Roentgen stereophotogrammetric analysis of the mallory head prosthesis without markers. *Acta Orthopaedica Scandinavica*, 67(3), 1996.

ER Valstar. *Digital roentgen stereophotogrammetry. Development, validation, and clinical application.* PhD thesis, 2001.

ER Valstar, FW de Jong, HA Vrooman, PM Rozing, and JHC Reiber. Model-based roentgen stereophotogrammetry of orthopaedic implants. *Journal of Biomechanics*, 34(6):715–722, 2001.

Everine B van de Kraats, Graeme P Penney, Dejan Tomazevic, Theo van Walsum, and Wiro J Niessen. Standardized evaluation methodology for 2-d-3-d registration. *Medical Imaging, IEEE Transactions on*, 24(9):1177–1189, 2005.

Andreas Wächter and Lorenz T Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming*, 106(1):25–57, 2006.

N Wolterbeek, EH Garling, B Mertens, ER Valstar, and RGHH Nelissen. Mobile bearing knee kinematics change over time. a fluoroscopic study in rheumatoid arthritis patients. *Clinical Biomechanics*, 24(5):441–445, 2009.

N Wolterbeek, EH Garling, BJA Mertens, HMJ van der Linden, RGHH Nelissen, and ER Valstar. Kinematics of a highly congruent mobile-bearing total knee prosthesis. *Knee Surgery, Sports Traumatology, Arthroscopy*, 20(12):2487–2493, 2012a.

N Wolterbeek, RGHH Nelissen, and ER Valstar. No differences in in vivo kinematics between six different types of knee prostheses. *Knee Surgery, Sports Traumatology, Arthroscopy*, 20(3):559–564, 2012b.

SJ Wright and J Nocedal. *Numerical optimization*, volume 2. Springer New York, 1999.

P. Wunsch and G. Hirzinger. Registration of cad-models to images by iterative inverse perspective matching. In *Pattern Recognition, 1996., Proceedings of the 13th International Conference on*, volume 1, pages 78–83. IEEE, 1996.

B.M. You, P. Siy, W. Anderst, and S. Tashman. In vivo measurement of 3-d skeletal kinematics from sequences of biplane radiographs: application to knee kinematics. *Medical Imaging, IEEE Transactions on*, 20(6):514–525, 2001.

D. Zhao, S.A. Banks, D.D. D'Lima, C.W. Colwell Jr, and B.J. Fregly. In vivo medial and lateral tibial loads during dynamic and high flexion activities. *Journal of Orthopaedic Research*, 25(5):593–602, 2007.

S. Zuffi, A. Leardini, F. Catani, S. Fantozzi, and A. Cappello. A model-based method for the reconstruction of total knee replacement kinematics. *Medical Imaging, IEEE Transactions on*, 18(10):981–991, 1999.

LIST OF PUBLICATIONS

A.H. Prins, B.L. Kaptein, B.C. Stoel, R.G.H.H. Nelissen, J.H.C. Reiber, E.R. Valstar (2008). Handling modular hip implants in model-based RSA. *Journal of Biomechanics* 41,(14), 2912-2917

A.H. Prins, B.L. Kaptein, B.C. Stoel, R.G.H.H. Nelissen, J.H.C. Reiber, E.R. Valstar (2011). Integrated contour detection and pose estimation for fluoroscopic analysis of knee implants. *Journal of Engineering in Medicine* 225,(8), 753-761

A.H. Prins, B.L. Kaptein, B.C. Stoel, J.H.C. Reiber, E.R. Valstar (2010). Detecting femur-insert collisions to improve precision of fluoroscopic knee arthroplasty analysis. *Journal of biomechanics* 43,(4), 694-700

A.H. Prins, B.L. Kaptein, S.A. Banks, B.C. Stoel, R.G.H.H. Nelissen, E.R. Valstar (2013). Detecting condylar contactloss using single-plane fluoroscopy: A comparison with *in vivo* forcedata and *in vitro* bi-plane data. *Journal of biomechanics 2014 47 (7), 1682–1688*

A.H. Prins, B.L. Kaptein, B.C. Stoel, D.J.P. Lahaye, E.R. Valstar (2014). Performance of Optimization in Single-plane Fluoroscopic Analysis for TKA. *Submitted to Journal of Biomechanics*

CURRICULUM

André Prins was born on October 7, 1981 in Assen, The Netherlands. In 1999 he graduated from the VWO "Het Stellingwerf College" in Oosterwolde and he started his study Computer Science at the University of Groningen. He specialized in "Computational Science and Visualization" and carried out his master's thesis work in the area of Computational Geometry. His master's thesis work "Certified reconstruction of convex curves using radial basis functions" was supervised by Gert Vegter and he graduated in December 2004.

Following his graduation, he worked for two years as a consultant for Shell in 2005 after which he joined the Department of Orthopaedics of the Leiden University Medical Center in 2006 as a PhD student. His PhD project was part of the European DeSSOS-project where the objective was to develop decision support software for orthopaedic surgery. Fluoroscopy and Roentgen Stereophotogrammetry were used for the validation of the knee models in the DeSSOS project and he investigated the reliability and usability of model-based shape matching for RSA and Fluoroscopy.

After finishing his PhD project he worked for three years as a senior software engineer at Motek Medical in Amsterdam. Afterwards he joined Alten as a scientific software engineer and is now working again as a consultant for Shell.

ACKNOWLEDGEMENTS

The research in this thesis has been done at the orthopaedic department of the Leiden University Medical Center and I thank my supervisors Edward, Bart and Berend for their support and patience while writing the articles and this thesis.

I have collaborated with several colleagues from both the orthopaedic department and from the LKEB and I thank all of them for the advice and support they have given me. Nienke deserves mentioning for actually using the software I developed and giving valuable feedback on it. I would like to thank Christiaan and Emiel for sharing the burden of completing a thesis. We have had many discussions about dealing with co-authors and the academic world in general.

I would like to thank my parents for supporting me in my studies all my career. I would also like to thank Hans and Elly for providing me with all the time and space to finish the thesis in all those visits to Berkenwoude.

Finally, I would like to thank Andrea for being with me and supporting me throughout the difficult parts near the end of the thesis.